ROBUSTNESS OF TWO MULTIVARIATE TESTS TO VARIANCE-COVARIANCE HETEROSCEDASTICITY AND NONNORMALITY WHEN TOTAL-SAMPLE-SIZE-TO-VARIABLE RATIO IS SMALL

BY

WEN-YING LIN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1991

ACKNOWLEDGMENTS

I would like to acknowledge my gratitude to the many individuals who have assisted me in completing this study. First, I would like to thank Dr. James J. Algina, chairman of my doctoral committee, for his direction, guidance, and assistance throughout all phases of this dissertation. His outstanding professional expertise and editorial skills have contributed invaluably to the accuracy and depth of the dissertation. Second, I want to extend my sincere appreciation to Dr. Robert R. Sherman for the confidence, encouragement, and advice that he never failed to provide whenever I needed them. He is not only a great teacher but also a precious friend. I am also grateful to the other members of my committee, Dr. Linda M. Crocker, Dr. M. David Miller, and Dr. Walter R. Cunningham, for their constructive suggestions and continuous support. Furthermore, heartfelt thanks are extended to my family for their deep love and enduring patience. It is to my family that this dissertation is dedicated. Last, but by no means the least, my humble gratitude is to my Lord Jesus who made it all possible. "Now to the King eternal, immortal, invisible, the only God, be honor and glory for ever and ever. Amen." (I Timothy 1:17)

TABLE OF CONTENTS

		page
ACKNOWLEDGMENTS		ii
ABSTRACT		v
CHAPTERS		
1	INTRODUCTION	1
	The Problem	4
	Purpose of the Study	4
	Significance of the Study	5
2	REVIEW OF LITERATURE	6
	Multivariate Analysis of Variance	6
	Hotelling's Two Sample T^2	8
	Robustness of T^2 under V-C Heterogeneity	9
	Robustness of T^2 under Nonnormality	11
	Robustness of T^2 under V-C Heterogeneity and Nonnormality	12
	The Behrens-Fisher Problem	13
	James' Two Sample Tests	14
	Yao's Test	15
	Johansen's K Sample Test	16
	Robustness of James' Tests	16
	Robustness of Yao's Test	18
	Robustness of Johansen's Test	20
	Nonparametric Procedure	21
	Rank Transformation Procedure	22
	Summary	23
3	METHODOLOGY	24
	Design	24
	Invariance Characteristic of the Test Statistic	29
	Simulation Procedure	31

4 RESULTS AND DISCUSSION..... 33

Frequency Distributions of ts..... 33
Significance Tests..... 50

5 CONCLUSIONS AND RECOMMENDATIONS..... 83

APPENDICES

A FREQUENCY DISTRIBUTIONS OF t WHEN $\alpha = .01$..... 91

B FREQUENCY DISTRIBUTIONS OF t WHEN $\alpha = .10$..... 103

C SUMMARY ANOVA TABLES FOR $\alpha = .01$ AND $\alpha = .10$..... 115

REFERENCES..... 122

BIOGRAPHICAL SKETCH..... 126

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

ROBUSTNESS OF TWO MULTIVARIATE TESTS TO VARIANCE-COVARIANCE HETEROSCEDASTICITY AND NONNORMALITY WHEN TOTAL-SAMPLE-SIZE-TO-VARIABLE RATIO IS SMALL

BY

Wen-ying Lin

May, 1991

Chairman: James J. Algina
Major Department: Foundations of Education

A total of 228 simulated experiments--192 with unequal sample sizes and 36 with equal sample sizes--were conducted to study the robustness of James' second order test and Johansen's test to variance-covariance (v-c) heterogeneity and nonnormality when total-sample-size-to-variable ratio (N/p) is small. The design is complete for the normal and beta distributions when $\Sigma_2 \neq d^2\Sigma_1$ but incomplete for the exponential distribution when $\Sigma_2 \neq d^2\Sigma_1$ and for the normal distribution when $\Sigma_2 = d^2\Sigma_1$.

The results of the study indicated that N/p and the type of heterogeneity of v-c matrices have a substantial effect on the error rates. In general, as N/p increases, the error rates decrease. With the normal distribution and $\Sigma_2 \neq d^2\Sigma_1$, the two tests generally produce satisfactory

error rates when N/p = 7 or 9. However, with the normal distribution and $\Sigma_2 = d^2\Sigma_1$, James' test is nonrobust for N/p = 5 or 7; and Johansen's test is nonrobust for all three values of N/p. Similarly, skewness and kurtosis also have a strong impact on the error rates. Specifically, the normal distribution results in error rates significantly closer to the nominal value than does the beta distribution, and in turn the beta distribution results in better error rates than does the exponential distribution. Further, the error rates are considerably larger under the negative condition than under either the positive condition or the equal-sample-size condition. Between the positive condition and the equal-sample-size condition, the former seems to produce slightly superior error rates relative to the latter.

In conclusion, James' test performs slightly better than Johansen's test. In terms of relationship between sample sizes and v-c matrices, the two tests perform similarly under the positive and the equal-sample-size conditions, but James' test performs slightly better than Johansen' test under the negative condition. In addition, as p increases, the performance of James' test improves and that of Johansen's test declines.

CHAPTER 1
INTRODUCTION

In educational research, it is often necessary to compare two groups of subjects on the means of several dependent variables. That is, p response variables are observed from k = 2 groups of sizes n_1 and n_2, and the hypothesis of equal mean vectors for the two groups is tested. Under the assumptions of variance-covariance (v-c) homogeneity and of normality, Hotelling's (1931) T^2 is the most powerful test of the hypothesis. However, in practice data are unlikely to meet these two assumptions.

Numerous studies have been conducted on the robustness of T^2 with respect to violations of v-c homogeneity and/or normality. Both theoretical and, mostly, empirical findings have indicated that under nonnormality T^2 (for two samples) is quite robust, but that under v-c heteroscedasticity T^2 may not be robust when $n_1 = n_2$ and is likely not robust when $n_1 \neq n_2$.

As a result, the Behrens-Fisher problem--a comparison of the means of two groups under v-c heteroscedasticity--has been a subject of research in multivariate statistics. Several alternatives to T^2, intended to produce better results for the Behrens-Fisher problem, have been proposed.

Among the alternatives are James' (1954) first and second order tests, Yao's (1965) test, and Johansen's (1980) test. Distinguished by their critical values, all four tests use the statistic

$$T_v^2 = (\bar{x}_1 - \bar{x}_2)'(\frac{S_1}{n_1} + \frac{S_2}{n_2})^{-1}(\bar{x}_1 - \bar{x}_2)$$

where $\bar{x}_i$ and S_i are the mean vector and v-c matrix for the ith sample, i = 1, 2.

Yao (1965) conducted a simulation study for p = 2 to estimate Type I error rates of James' first order test and his own test. His results indicated that whereas both tests are quite robust under v-c heteroscedasticity, Yao's test is superior to James' test. Algina and Tang (1988) conducted a more extensive study than Yao's on the performance of T^2, Yao's test, and James' first order test. The ratios of total sample size to number of variables (N/p) used in Algina and Tang were 6, 10, and 20. They confirmed the superiority of Yao's test. Further, they found that when 10 $\leq$ N/p $\leq$ 20, Yao's test produces appropriate error rates for conditions in which p $\leq$ 10, $n_1:n_2 \leq 2:1$, and d $\leq$ 3; when N/p = 20, Yao's test produces appropriate error rates when $n_1:n_2 \leq 5:1$ and d $\leq$ 3. This was true both for the case in which one v-c matrix was a multiple of the second v-c matrix ($\Sigma_i = d^2\Sigma_i$) and in cases with more complicated heteroscedasticity. However, when N/p = 6, they found Yao's

test to be liberal. In this condition they studied only the case in which $\Sigma_{i'} = d^2\Sigma_i$.

Guided by the recommendations of Algina and Tang (1988), later studies on these alternatives to Hotelling's T^2 have avoided using N/p ratios that are smaller than 10. Tang (1989) investigated the robustness of four multivariate tests for k = 3 groups when the v-c matrices are heteroscedastic. The N/p ratios used were 10, 15, and 20. In addition to finding that James' second order and Johansen's tests perform better than James' first order test, Tang found that the ratio of N to p has a strong impact on the performance of the tests. In general, as the ratio increases, the tests become more robust.

Algina and Oshima (1990b) investigated the robustness of Yao's, James' first and second order, and Johansen's tests for k = 2 groups under v-c heterogeneity and nonnormality. Values of 10 and 20 were used for the N/p ratio. Their study indicated that these four alternatives to Hotelling's T^2 may not be robust when the sampled distributions are skewed and have positive kurtosis, and the v-c matrices are heteroscedastic. In particular, the tests are seriously nonrobust with lognormal and exponential distributions. By comparison, Yao's, James' second order, and Johansen's tests tend to perform better than James' first order test. Hence, they did not recommend the first order test for use.

The Problem

For k = 2, it is known that James' first-order, James' second-order, Yao's, and Johansen's tests have Type I error rates near the nominal level provided that $N/p \geq 10$ and the data are not too skewed. Moreover, the last three tests perform better than the first test. Little is known about the performance of the tests when $N/p < 10$. For $N/p = 6$, multivariate normal distributions, and $\Sigma_i' = d^2\Sigma_i$, James' first order test and Yao's test are liberal (Algina & Tang, 1988). The performance of the tests for $6 < N/p < 10$ has not been investigated either for $\Sigma_i' = d^2\Sigma_i$ or for more complicated forms of heteroscedasticity. In addition, it is not known whether the effect of skewness is more severe for $N/p < 10$ than it is for $N/p \geq 10$.

Both James' second order and Johansen's tests can be generalized to more than two groups, whereas Yao's test cannot. Previous studies have indicated that Yao's and Johansen's tests produce very similar results. Moreover, in theory, James' second order test should perform better than Johansen's test when sample sizes are small. Thus, only James' second order and Johansen's tests were considered in the study.

Purpose of the Study

The purpose of this study was to investigate for k = 2 groups the robustness of James' second order and Johansen's

tests to v-c heteroscedasticity and nonnormality when the ratio of total sample size to number of variables is small.

Significance of the Study

It is common in educational research to test hypotheses on mean vectors for two or more populations. When data are from moderately skewed distributions, both James' second order and Johansen's tests will produce appropriate Type I error rates under v-c heteroscedasticity, if researchers can follow the recommendations to have $N/p \geq 10$ and preferably $N/p \geq 20$. In practice, large sample sizes are sometimes impractical and, therefore, large values of N/p are not possible. For instance, when human subjects are involved in research, substantial sample sizes may not be available and, even if available, are often costly. Thus, investigating the performance of these alternatives to Hotelling's T^2 for small values of N/p is merited.

CHAPTER 2
REVIEW OF LITERATURE

Multivariate Analysis of Variance

Situations that involve testing the equality of the mean vectors of several ($k \geq 2$) populations are often encountered in educational research. In these situations, multivariate analysis of variance (MANOVA) can be used to analyze the data. MANOVA requires the following assumptions:

1. All k populations have a common v-c matrix Σ.
2. Each of the k populations is multivariate normal.
3. The random samples from different populations are independent.

Although no single MANOVA invariant test is uniformly most powerful, four MANOVA tests are commonly used:

1. Roy's largest root;
2. Hotelling-Lawley trace;
3. Wilks' likelihood ratio;
4. Pillai-Bartlett trace.

The v-c homogeneity and normality assumptions in MANOVA may impose restrictions on its applicability. To investigate this possibility, Olson (1974) conducted a Monte Carlo study of the performance of six MANOVA tests,

including the preceding four, when these two assumptions are violated. Based on his study, Olson recommended the Pillai-Bartlett trace for its robustness. However, the actual Type I error rate of this test is still somewhat higher than its nominal error rate.

Using normal and nonnormal data, Pavur and Nath (1989) compared Type I error rates of Wilks' lambda with the Puri and Sen (1971) statistics both calculated using ranks and normal scores. Both chi-square and F approximations to the sampling distributions were investigated. Exponential, uniform, Cauchy, and lognormal distributions were selected to simulate nonnormal data. They concluded that in both normal and nonnormal homoscedastic situations, Wilks' lambda statistic computed from normal scores and used with the F approximation is more robust than the Puri and Sen tests. Also, the F approximation for the Puri and Sen statistic is more robust than the chi-square approximation.

Tang (1989) investigated Type I error rates of the Pillai-Bartlett trace, Johansen's test, James' first order test, and James' second order test when the v-c matrices are heteroscedastic. She found that the Pillai-Bartlett trace can be safely used only when the sample sizes are equal and the degree of v-c heteroscedasticity is small. In general, Johansen's and James' second order tests perform better than the Pillai-Bartlett trace and James's first order test. With respect to their significance levels, Johansen's test

tends to be liberal and James' second order test tends to be conservative. However, the Type I error rates for both tests are reasonably close to α.

Hotelling's Two Sample T^2

When two populations are involved, MANOVA reduces to Hotelling's (1931) T^2. Let the two random samples be denoted by

$$\{X_{ij},\ i = 1,\ 2;\ j = 1,\ 2,\ \ldots,\ n_i\}.$$

Then the mean vectors and v-c matrices for the samples are

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} ,$$

and

$$S_i = \frac{1}{n_i - 1}\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)' .$$

A natural generalization of the univariate t^2, the T^2 statistic is

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)' S_p^{-1} (\bar{x}_1 - \bar{x}_2)$$

where

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2} .$$

The quantity

$$F_{TSQ} = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} T^2$$

is an F-variate with p and $n_1 + n_2 - p - 1$ degrees of freedom. The null hypothesis is rejected at level α whenever

$$F_{TSQ} > F_{\alpha,p,n_1+n_2-p-1} .$$

Robustness of T^2 under V-C Heterogeneity

Investigations of the robustness of Hotelling's T^2 with respect to violation of v-c homogeneity have been conducted from both analytical (Ito & Schull, 1964) and empirical (Algina & Oshima, 1990a; Hakstian, Roed, & Lind, 1979; Holloway & Dunn, 1967; Hopkins & Clay, 1963) standpoints.

Ito and Schull (1964) studied the asymptotic properties of T^2 when $\Sigma_1 \neq \Sigma_2$. They found that for p = 1, 2, 3, and 4, the violation of v-c homogeneity does not substantially affect the significance level when the sample sizes are equal, but does affect the significance level seriously when they are not equal. Specifically, their study indicated that when $\Sigma_i = d^2\Sigma_j$ the actual Type I error rate (τ) is larger than the nominal Type I error rate (α) when the relationship between the sample sizes and v-c matrices is negative and τ is smaller than α when the relationship is positive. For example, for p = 2, $n_1:n_2$ = 2:1, and α = 0.05, τ = 0.091 when $\Sigma_2 = 2\Sigma_1$ and τ = 0.023 when $\Sigma_1 = 2\Sigma_2$.

With more complicated patterns of heteroscedasticity the discrepancy between τ and α has a complex relationship to variation in n_j and θ_j, the eigenvalues of $\Sigma_1\Sigma_2^{-1}$.

Hopkins and Clay (1963) also found that bivariate T^2 is quite robust under v-c heteroscedasticity when $n_1 = n_2 \geq 10$, but this robustness does not extend to unequal sample sizes. For example, for $\Sigma_2 = 1.6^2\Sigma_1$ and $\alpha = 0.05$, the estimated Type I error rate (τ) is 0.052 when $n_1 = n_2 = 10$, but τ is 0.019 when $n_1 = 10$ and $n_2 = 20$. Similarly, Hakstian, Roed, and Lind (1979) showed that T^2 is generally robust when $n_1 = n_2$. However, when $n_1 \neq n_2$, T^2 becomes increasingly less robust as p and v-c matrix heterogeneity increase. Specifically, when $n_1{:}n_2 = 2{:}1$, τ departs only moderately from α for $p = 2$ and mild v-c matrix heterogeneity, but τ departs markedly from α for $p = 10$ and moderate to substantial v-c matrix heterogeneity. For $n_1 \neq n_2$ their findings were consistent with those of Ito and Schull (1964): T^2 is conservative when the larger sample has the larger v-c matrix and liberal otherwise. Holloway and Dunn (1967) found that for both equal and unequal sample sizes the discrepancy between τ and α increases as p increases, or as the sample size decreases. With sample sizes unequal and in small ratios, Algina and Oshima (1990a) found that T^2 can be seriously nonrobust, even with relatively mild v-c heteroscedasticity and small values of p. For example, when $\Sigma_2 = 1.5^2\Sigma_1$ and $p = 2$, a

sample-size ratio as small as 1.1:1 can produce unacceptable Type I error rates.

In sum, when $n_1 = n_2$, factors such as sample sizes, number of variables, and the degree of v-c heterogeneity may affect the robustness of T^2. On the other hand, when $n_1 \neq n_2$, T^2 leads to unacceptable Type I error rates. In particular, T^2 often results in underestimation of significance when the relationship between the sample sizes and v-c matrices is positive and overestimation when the relationship is negative. This discrepancy becomes more pronounced as the number of variables, the degree of v-c heterogeneity, and the magnitude of the inequality of two samples increase.

Robustness of T^2 under Nonnormality

The effect of nonnormality on Hotelling's T^2 for two samples has been studied analytically by Ito (1969) and empirically by Everitt (1979) and Nath and Duran (1983). As long as sample sizes are very large, Ito concluded that the behavior of T^2 is not seriously affected by violation of the normality assumption. Everitt conducted a Monte Carlo study on T^2 in which p = 2, 4, 6, or 10 variables were considered. Uniform, exponential, and lognormal distributions were used to simulate nonnormal data. His results showed that departures from normality due to skewness are likely to lead to moderately conservative tests. For example, when samples were from the very skewed lognormal distribution, α was

typically between .03 and .04 for $\alpha = .05$. However, when samples were from the non-skew uniform distribution, the observed levels were very close to the nominal levels. Nath and Duran studied the robustness of T^2 when $p = 2$. Uniform, exponential, lognormal, and Cauchy distributions were used to generate nonnormal observations. Their results are in close agreement with those of Everitt. That is, T^2 is fairly robust with respect to uniform and exponential distributions but somewhat conservative with respect to the lognormal distribution. In addition, for the heavy-tailed Cauchy distribution, T^2 tends to be overly conservative.

Robustness of T^2 under V-C Heterogeneity and Nonnormality

Zwick (1986) and Algina and Oshima (1990b) conducted simulation studies on the robustness of T^2 under v-c heterogeneity and nonnormality. Zwick simulated data from a beta distribution with skewness -.72 and kurtosis .31 and concluded that the effect of nonnormality on T^2 is slight, but that of v-c heteroscedasticity on T^2 is significant, especially when sample sizes are unequal. For example, when $n_1 \neq n_2$ and $\alpha = 0.01$, $\hat{\alpha} = 0.011$ under v-c homogeneity and nonnormality, $\hat{\alpha} = 0.002$ under v-c heterogeneity and normality, and $\hat{\alpha} = 0.001$ under v-c heterogeneity and nonnormality. Further, Zwick's results are consistent with those of the preceding studies in that T^2 tends to be conservative when a larger sample is paired with a larger v-

c matrix, and is liberal otherwise, regardless of whether the data are sampled from normal distributions.

Generating nonnormal observations from six types of distributions, Algina and Oshima (1990b) found that T^2 is seriously nonrobust with heteroscedastic v-c matrices and unequal sample sizes. For example, with the lognormal distribution and in the negative condition when $\alpha = .05$, $N/p = 10$, $p = 2$, and $n_1:n_2 = 2:1$, the estimated Type I error rate for T^2 was .149. In fact, under extreme heteroscedasticity, T^2 is nonrobust even when sample sizes are equal.

The Behrens-Fisher Problem

The lack of robustness to heteroscedasticity for T^2 calls for other alternatives to solve the heteroscedasticity problem. As a result, the Behrens-Fisher problem--a comparison of the means of two populations under v-c heteroscedasticity--has been a popular subject in multivariate research. Over the decades, many statisticians have investigated this topic using different methods and under various, more or less restrictive conditions.

The univariate version of the problem was first studied by Behrens (Yao, 1965), and his solution was formulated by Fisher (1935, 1939) in terms of fiducial probability. Welch (1938) studied the univariate Behrens-Fisher problem and derived an approximate degrees of freedom (APDF) solution. The test statistic he used is distributed approximately as

Student's t with degrees of freedom suitably adjusted. Later, Welch (1947) developed a series solution.

In the multivariate extension of the Behrens-Fisher problem, James (1954) extended the Welch series solution, Yao (1965) the Welch APDF solution to two samples, and Johansen (1980) the Welch APDF solution to k samples. When two samples are involved, all three used the statistic

$$T_v^2 = (\bar{x}_1 - \bar{x}_2)'\left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1}(\bar{x}_1 - \bar{x}_2)$$

which is asymptotically distributed as chi-square with p degrees of freedom.

James' Two Sample Tests

James' (1954) test is a multivariate generalization of his series solution (James, 1951) to the k-sample univariate Behrens-Fisher problem. For two samples, the critical value of James' first order test is

$$h_1(S_1, S_2; \alpha) = \chi_\alpha^2(p)\left[1 + \frac{1}{2}\left(\frac{k_1}{p} + \frac{k_2\chi_\alpha^2(p)}{p(p+2)}\right)\right],$$

where

$$k_1 = \sum_{i=1}^{2} \frac{[tr(I - W^{-1}W_i)]^2}{n_i - 1},$$

and

$$k_2 = 2\sum_{i=1}^{2} \frac{tr(I - W^{-1}W_i)^2}{n_i - 1} + \sum_{i=1}^{2} \frac{[tr(I - W^{-1}W_i)]^2}{n_i - 1}.$$

In these expressions, $W_i = (S_i/n_i)^{-1}$, $W = W_1 + W_2$, tr is the trace operator, and $\chi_\alpha^2(p)$ is the $(1 - \alpha)$th percentile of the chi-square distribution with p degrees of freedom. The critical value of James' second order test is

$$h_2(S_1,S_2;\alpha) = h_1(S_1,S_2;\alpha) + O[(n_i - 1)^{-2}]$$

where $O[(n_i - 1)^{-2}]$ are given in formula 6.7 of James' (1954) paper. The discrepancy between τ and α is of order $(n_i - 1)^{-1}$ for the first order test and of order $(n_i - 1)^{-2}$ for the second order test. Consequently, the second order test should be particularly advantageous with small sample sizes.

Yao's Test

Yao (1965) extended Welch's (1938, 1951) solution to the multivariate Behrens-Fisher problem. His critical value is $T^2(p,f;\alpha)$, the $(1 - \alpha)$th percentile of the distribution of Hotelling's T^2. Let $E_i = S_i/n_i$ and $E = E_1 + E_2$. Define $V_i = (\bar{X}_1 - \bar{X}_2)'E^{-1}E_iE^{-1}(\bar{X}_1 - \bar{X}_2)$. The APDF f is defined by

$$\frac{1}{f} = \sum_{i=1}^{2} \frac{1}{(n_i - 1)} \left(\frac{V_i}{T_v^2}\right)^2 .$$

The transformation

$$F_v = \frac{f - p + 1}{pf} T_v^2$$

can also be used. Then the critical region is

$$F_v > F_\alpha(p,f-p+1).$$

Johansen's K Sample Test

Johansen (1980) generalized Welch's (1951) APDF solution to deal with the multivariate k-sample Behrens-Fisher problem. His critical value is $CF_{\alpha}(p,q)$, where

$$C = p + 2A - 6A/(p + 2),$$

$$q = p(p + 2)/3A,$$

and

$$A = \sum_{i=1}^{2} \frac{tr(I - W^{-1}W_i)^2}{2(n_i - 1)} + \sum_{i=1}^{2} \frac{[tr(I - W^{-1}W_i)]^2}{(n_i - 1)} .$$

Robustness of James' Tests

Yao (1965) conducted a simulation study of Type I error rates of James' first order test and his own test for p = 2. His results showed that the two tests are quite robust under v-c heteroscedasticity. A more extensive study of the behavior of T^2, James' first order test, and Yao's test was conducted by Algina and Tang (1988). In terms of control of Type I error rates, both James' and Yao's tests perform better than T^2. Neither test tends to be conservative, with the estimated error rates of James' test being larger than those of Yao's.

Tang (1989) studied the robustness of James' first and second order tests for k = 3 groups when the v-c matrices are heteroscedastic. The N/p ratios she used were 10, 15, and 20. She found that N/p has the dominant effect on the

tests. As N/p increases from 10 to 20, the performance of both tests improves. For example, with one pattern of heteroscedasticity, $\alpha = .05$, and $n_1:n_2:n_3 = 6:12:12$, the estimated Type I error rate decreases from .11 to .061 for the first order test and from .059 to .045 for the second order test when N/p increases from 10 to 20. On the other hand, as the degree of heteroscedasticity increases, the performance of both tests generally declines. Contrary to the performance of the first order test, the second order test performs better in the negative condition than in the positive condition. In addition, Tang found that the change in p has a substantial effect on the performance of the second order test. Specifically, the test is more robust for $p = 3$ than for $p = 6$. In terms of controlling Type I error rates, the first order test performs poorly under all conditions, especially when the sample-size ratio is in the extreme. Hence, Tang did not recommend James' first order test for use. In contrast, the second order test performs well in most conditions, even when the sample-size ratio is extreme. For example, in one negative condition with $\alpha = .05$, $N/p = 20$, and $p = 6$, the estimated error rate is .076 for the first order test and .049 for the second order test when $n_1:n_2:n_3 = 24:48:48$. Furthermore, Tang concluded that the second order test can be used when $N/p \geq 10$.

Algina and Oshima (1990) studied the effect of v-c heteroscedasticity and nonnormality upon James' first and

second order tests. Uniform, exponential, t(5), beta (5, 1.5), Laplace, and lognormal distributions were used to generate nonnormal data. Their study showed that both tests may not be robust when the v-c matrices are heteroscedastic and when the sampled populations are skewed and have positive kurtosis. Specifically, the two tests are seriously nonrobust with exponential (skewness = 2 and kurtosis = 6) and lognormal (skewness = 6.18 and kurtosis = 110.93) distributions and slightly nonrobust with the beta distribution (skewness = -0.82 and kurtosis = 0.28). For example, in the negative condition when $\alpha = .05$, $p = 10$, $N/p = 20$, and $n_1:n_2 = 4:1$, the estimated error rates for the first order test and the second order test are respectively .111 and .095 under exponential distribution, .221 and .197 under lognormal distribution, and .083 and .063 under beta distribution. They are fairly robust with respect to the remaining three symmetric distributions. James' second order test is slightly superior to the first order test.

Robustness of Yao's Test

Yao's (1965) simulation study showed that both his test and James' first order test are quite robust, but his test is slightly superior to James' test. The generality of his results is limited, since only the bivariate case was considered.

Algina and Tang (1988) extended Yao's study to investigate the robustness to heteroscedasticity of James'

first order and Yao's tests. The N/p ratios used were 6, 10, and 20. They concluded that Yao's test produces better results than James' test. In addition, they concluded that Yao's test can be used safely if $p \leq 10$, $N/p \geq 10$, and the ratio of the larger to the smaller sample size is 2:1 or smaller. Also, they pointed out that if $N/p \geq 20$, Yao's test can be used safely if $p = 2$ and the ratio is 5:1 or smaller; if $p = 6$ and the ratio is 3:1 or smaller; and if $p = 10$ and the ratio is 4:1 or smaller. However, when $N/p = 6$, Yao's test is liberal in most cases.

Algina and Oshima (1990b) studied the robustness of Yao's test under v-c heteroscedasticity and nonnormality for $k = 2$. Guided by the recommendations of Algina and Tang (1988) for the safe use of Yao's test, they used 10 and 20 for the ratio N/p. Their study indicated that when the sampled distributions are symmetric, Yao's test is quite robust under v-c heteroscedasticity and unequal sample sizes. However, when the distributions are skewed and have positive kurtosis, the test tends to be liberal. For example, in the negative condition when $\alpha = 0.05$, $p = 2$, $n_1:n_2 = 2:1$, and $N/p = 10$, the estimated error rate of the test is .122 under the lognormal distribution. Their findings are consistent with those of Clinch and Keselman (1982), which indicated that the univariate counterpart to Yao's test is liberal with skewed distribution.

Robustness of Johansen's Test

Tang (1989) estimated Type I error rates of Johansen's test for k = 3 groups under v-c heterogeneity. She found that, like James' first order test, Johansen's test performs better in the positive condition than in the negative condition. For example, when $\alpha = .05$, $N/p = 15$, $p = 6$, and $n_1:n_2:n_3 = 22:22:46$, the estimated Type I error rate is .0585 under the positive condition and .071 under the negative condition. Again, similar to the performance of James' first order test, the performance of the Johansen's test declines when the sample-size ratio increases. In general, Johansen's test performs well, though it tends to be slightly liberal. Tang recommended use of the test in situations (a) where sample sizes are equal and (b) where sample sizes are unequal, and $N/p \geq 15$.

Estimating Type I error rates for two groups under v-c heterogeneity and nonnormality, Algina and Oshima (1990b) found that Johansen's test suffers the same inadequacy as that of James' and Yao's tests. That is, under v-c heterogeneity and with unequal sample sizes, the test may not be robust when data are from skewed distributions with positive kurtosis. However, along with James' second order and Yao's tests, Johansen's test produces appropriate error rates when the distributions are symmetric. Though both James' second order and Johansen's tests are similar in

performance, Algina and Oshima were slightly in favor of the latter because of its simplicity in computations.

Nonparametric Procedure

In practice, populations do not always satisfy the normality assumption of the Behrens-Fisher problem. Other procedures whose validity does not depend on such an assumption are needed.

Fligner and Policello (1981) proposed a method for modifying many of the standard nonparametric rank tests for the two-sample location problem. After modification, the procedures can be used to test for equality of the medians with fewer assumptions on the shapes of the underlying populations. The modified statistics are still distribution-free when the populations are symmetric and identical. In addition, when the populations have common medians but different shapes, the modified statistics are asymptotically distribution-free as long as the populations are symmetric.

With fairly small and unequal sample sizes, Fligner and Policello (1981) conducted a Monte Carlo study on the robustness of the pooled variance t test, the Welch t test (1937, 1947), the Mann-Whitney-Wilcoxon (MWW; Mann and Whitney, 1947; Wilcoxon, 1945) test, and the modified MWW test. Four symmetric distributions--uniform, contaminated normal, Laplace, and Cauchy--were selected to generate nonnormal data. Their results showed that the modified MWW

test performs better than the other three tests in most situations and its Type I error rates are reasonably close to α. However, this approach has one serious limitation. That is, it is valid only for symmetric distributions.

Rank Transformation Procedure

As a way of dealing with nonnormal data, Conover and Iman (1981) developed a technique, called the rank transformation approach, to serve as a bridge between parametric and nonparametric statistics. That is, the usual parametric procedures are applied to the ranks or normal scores instead of to the original data.

To study the effectiveness of the rank transformation approach, Nath and Duran (1983) conducted a simulation study for p = 2 to assess the robustness of Hotelling's T^2, rank T^2, and Puri and Sen's rank statistic under nonnormality. Uniform, Cauchy, exponential, and lognormal distributions were used to generate nonnormal data. Their results showed that the rank T^2 is robust across all α levels, distributions, and sample sizes except $(n_1,n_2) = (7,3)$.

Using the rank transformation approach, Zwick (1986) compared Type I error rates of the rank and normal scores procedures to those of T^2 under v-c heteroscedasticity and nonnormality when p = 5. Hotelling's T^2 was calculated using ranks and normal scores and substituted into the test statistic

$$\chi^2 = \frac{(N - 1)T^2}{N - 2 + T^2}$$

where χ^2 is asymptotically distributed as chi-square with $p(k - 1)$ degrees of freedom. A beta (9,2.5) distribution, with skewness and kurtosis are -0.72 and 0.31, respectively, was used to simulate nonnormal data. She found that the rank and normal scores procedures, though better than T^2, still produced unsatisfactory results. For example, in the case of unequal variances and negative pairing between v-c matrices and sample sizes, both procedures produced error rates far in excess of α. Overall, her study was limited because only a small number of conditions were considered.

Summary

Based on the preceding review, it can be concluded that James' second order test, Yao's, and Johansen's tests tend to produce appropriate results for the Behrens-Fisher problem when the sampled populations are symmetric and $N/p \geq 10$. However, except for the study by Algina and Tang (1988), which included 6 as one of the values for N/p, none of the previous studies took values smaller than 10 into consideration. Thus, without further investigation of Type I error rates for smaller values of N/p, the recommendation for using at least 10 for N/p may be unnecessarily restrictive.

CHAPTER 3
METHODOLOGY

In this chapter, the design and simulation procedure used for the study are described. The design was based on a review of relevant literature and on the consideration that experimental conditions used in the simulation should be similar and applicable to practical educational research.

Design

The factors considered in the study were type of distribution, number of variables (p), sample-size ratio (n_1:n_2), total-sample-size-to-variable ratio (N/p), degree and type of heterogeneity of v-c matrices, and significance level. The levels of these factors are given in the following paragraphs.

Type of distribution. Three types of distribution--beta (2.5, 1), exponential, and normal--were included in this study. Algina and Oshima (1990b) showed that when the distributions are beta (5, 1.5), exponential, and lognormal, the four tests are not robust under v-c heteroscedasticity. The lack of robustness is most severe for the lognormal distribution (skewness = 6.18 and kurtosis = 110.93) and least severe for the beta distribution (skewness = -0.82 and kurtosis = 0.28).

The skewness and kurtosis of real-life data are often not as extreme as the skewness and kurtosis of those from lognormal distribution. Hence, in addition to the normal distribution, this study included only the less skewed beta and exponential (skewness = 2.0 and kurtosis = 6.0) distributions for consideration.

In educational research, data, such as from minimum competency tests or advanced placement tests, are often from moderately skewed distributions with negative kurtosis (i.e., short tail) because of floor or ceiling effects. Pearson and Please (1975) suggested that robustness studies should focus on distributions with skewness and kurtosis having absolute values less than 0.8 and 0.6, respectively. Thus, the chosen beta (2.5, 1) distribution, whose coefficients of skewness and kurtosis are respectively -0.73 and -0.24, meets all the above conditions.

Number of variables (p). Data were generated using p = 2, 6, and 10 variables. These same values of p were also used by both Algina and Tang (1988) and Algina and Oshima (1990b). Values of p that are larger than 10 are rarely encountered in educational research. As such, p = 10 was chosen as the upper limit for the study.

Sample-size ratio (n_1:n_2). Three ratios, n_1:n_2 = 1:1, 1.5:1, and 2:1, were used. In Algina and Tang's (1988) study, some of the sample-size ratios were fairly large. For instance, the most extreme ratio used was 5:1. However,

in experimental studies and in field studies, it is common to have sample-size ratios between 1:1 and 2:1. Further, Algina and Oshima (1990b) pointed out that the error rates increase as the sample-size ratio increases. Consequently, 2:1 was chosen as the largest sample-size ratio. Due to fixed total sample size, these three ratios needed slight adjustments in most cases.

Total-sample-size-to-variable ratio (N/p). The ratios N/p were 5, 7, and 9. Except for the study by Algina and Tang (1988), which included 6 as one of the values for N/p, all later studies have avoided using N/p that are smaller than 10. However, in many practical situations, large sample sizes are often not possible because of time and money constraints. Thus, only smaller values of the ratio were used in this study.

With the ratios N/p = 5, 7, and 9, the total sample sizes were 5p, 7p, and 9p. The sample sizes for k = 2 groups used in this study are shown in Tables 1, 2, and 3.

Degree and type of heterogeneity of v-c matrices. Data for group 1 were generated from a p x p identity v-c matrix I_p and data for group 2 from a p x p diagonal v-c matrix D^2. The first p/2 diagonal elements of D^2 were d^2 and the other p/2 diagonal elements were 1. Both the positive and negative relationships between sample size and v-c matrices were investigated.

Table 1

Sample Sizes When p = 2

N/p	n_1	n_2
5	5	5
5	6	4
5	7	3
7	7	7
7	8	6
7	9	5
9	9	9
9	11	7
9	12	6

Table 2

Sample Sizes When p = 6

N/p	n_1	n_2
5	15	15
5	18	12
5	20	10
7	21	21
7	25	17
7	28	14
9	27	27
9	32	22
9	36	18

Table 3

Sample Sizes When p = 10

N/p	n_1	n_2
5	25	25
5	30	20
5	33	17
7	35	35
7	42	28
7	47	23
9	45	45
9	54	36
9	60	30

Two levels of d, 1.5 and 3.0, were used to simulate the degree of heterogeneity of v-c matrices. Algina and Tang (1988) studied cases with d = 1.5, 2.0, 2.5, and 3.0; Tang (1989) with d = $1.5^{1/2}$ and 3.0; Algina and Oshima (1990b) with d = 1.5 and 3.0. Therefore, the two levels selected for this study represent moderate to larger degrees of heterogeneity investigated in other studies.

An additional set of data was generated from the normal distribution with $\Sigma_2 = d^2\Sigma_1$ to study Type I error rates under the negative condition. Specifically, data for group 1 were simulated from a p x p identity v-c matrix and data for group 2 from a p x p diagonal v-c matrix with diagonal elements all equal to d^2.

Significance level. Three levels of significance were used in this study. They were .01, .05, and .10.

The present design is incomplete for the exponential distribution and for the normal distribution when $\Sigma_2 = d^2\Sigma_1$. Initial runs of 12 negative conditions for the exponential distribution indicated that both James' second order and Johansen's tests are seriously nonrobust when N/p = 9. As such, further runs of the remaining experimental conditions for this distribution are not merited.

Cases where $\Sigma_2 = d^2\Sigma_1$ are not often found in real-life data. Therefore, in these cases, only the 36 negative conditions for the normal distribution were run to study the impact of type of heterogeneity on Type I error rates.

With the above restrictions taken into consideration, this design consisted of 192 experimental conditions with unequal sample sizes and 36 with equal sample sizes. Hence, a total of 228 experimental conditions were obtained.

Invariance Characteristic of the Test Statistic

Let Σ_1 and Σ_2 be the v-c matrices of size p x p for populations 1 and 2, respectively. Since both Σ_1 and Σ_2 are positive definite, there exists a p x p nonsingular matrix C such that $C'\Sigma_1 C = I_p$ and $C'\Sigma_2 C = D^2$, where I_p is a p x p identity matrix and D^2 is a p x p diagonal matrix. As shown below, the test statistic T_v^2 for both James' second order test and Johansen's test is invariant with respect to nonsingular transformations $Y_{ij} = CX_{ij}$, i = 1, 2 and j = 1,

2, ..., n_i, where X_{ij} is distributed with zero mean vector and v-c matrix = I_p and X_{2j} with zero mean vector and v-c matrix = D^2.

Suppose C is a p x p nonsingular matrix, let $Y_{ij} = CX_{ij}$ and so $\bar{Y}_i = C\bar{X}_i$. Then

$$S_1^* = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)(Y_{1j} - \bar{Y}_1)' .$$

Substituting CX_{1j} for Y_{1j} and $C\bar{X}_1$ for $\bar{Y}_1$, we have

$$S_1^* = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1} (CX_{1j} - C\bar{X}_1)(CX_{1j} - C\bar{X}_1)' .$$

Then

$$S_1^* = C[\frac{1}{n_1 - 1}\sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)(X_{1j} - \bar{X}_1)']C' .$$

Hence, we have $S_1^* = CS_1C'$. Similarly, $S_2^* = CS_2C'$.

Now T_v^{2*} computed from the Y_{ij}'s is

$$T_v^{2*} = (\bar{Y}_1 - \bar{Y}_2)'(\frac{S_1^*}{n_1} + \frac{S_2^*}{n_2})^{-1}(\bar{Y}_1 - \bar{Y}_2) .$$

By substitutions, we obtain

$$T_v^{2*} = (C\bar{X}_1 - C\bar{X}_2)'(\frac{CS_1C'}{n_1} + \frac{CS_2C'}{n_2})^{-1}(C\bar{X}_1 - C\bar{X}_2) .$$

By some mathematical manipulations, we have

$$T_v^{2*} = (\bar{X}_1 - \bar{X}_2)'C'^{-1}(\frac{S_1}{n_1} + \frac{S_2}{n_2})^{-1}C^{-1}C(\bar{X}_1 - \bar{X}_2) .$$

Hence

$$T_v^{2*} = (\bar{X}_1 - \bar{X}_2)'(\frac{S_1}{n_1} + \frac{S_2}{n_2})^{-1}(\bar{X}_1 - \bar{X}_2) = T_v^2 .$$

Therefore, the test statistic is invariant to nonsingular transformation. Hence, there is no loss of generality by simulating data only from an identity v-c matrix for group 1 and from a diagonal v-c matrix for group 2.

Simulation Procedure

The simulation was conducted as 228 separate runs, one for each condition, with 2000 replications per condition. For each replication, the computations were done according to the following outline:

1. For each of the k = 2 groups, n_i x p (i = 1, 2) uncorrelated pseudo-random observations were generated from the target distribution--beta (2.5, 1), exponential, or normal (0, 1). A set of p of these observations constituted the data for a given case. Using the population expected value and standard deviation, the random observations on each of the p variates were standardized. Thus, within each group, all the p variates were identically distributed with zero mean and unit variance and with all intercorrelations among the p variates equal to zero.

2. Each p x 1 vector X_{2j} of observations in group 2 was premultiplied by an appropriate D to simulate v-c heteroscedasticity.

The use of uncorrelated observations is justified by the fact that T_v^2 and its corresponding critical values are invariant with respect to nonsingular transformations of the data. For each replication, the data were analyzed using James' second order and Johansen's tests. The proportion of 2000 replications that yielded significant results at α = .01, .05, and .1 were recorded.

CHAPTER 4
RESULTS AND DISCUSSION

This chapter is divided into two parts: the frequency distributions of estimated Type I error rates ($\hat{\alpha}$) for each of the three significance levels (α) are described in the first part; statistical analyses of the Type I error rates of the two tests are reported in the second part.

Frequency Distributions of $\hat{\alpha}$s

In general, the major patterns of the generated data of Type I error rates are similar for all three levels of significance. Hence, frequency distributions of $\hat{\alpha}$ will be presented only for $\alpha = .05$. Frequency distributions of $\hat{\alpha}$ for $\alpha = .01$ and .10 are given in Appendices A and B.

Frequency distributions of $\hat{\alpha}$s when $\alpha = .05$. Frequency distributions of $\hat{\alpha}$ are reported, in terms of N/p, in Figure 1 for the normal distribution and $\Sigma_2 \neq d^2\Sigma_1$. The mean and standard deviation of $\hat{\alpha}$ are .057 and .011 for Johansen's test and .055 and .009 for James' test. Further, for Johansen's test 30% of the $\hat{\alpha}$s are above the upper limit of the interval $\alpha \pm 2[\alpha(1-\alpha)/2000]^{1/2} = (.0403, .0597)$; for James' test 22% are above the upper limit. For both tests, the majority of the $\hat{\alpha}$s above the upper limit occur in conditions in which N/p = 5. For N/p = 5, 20% and 13% of

Figure 1

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .05$

Johansen's test		James' test
555	≥.085	5
5	≥.080	5
55	≥.075	55
755	≥.070	
997555	≥.065	5559
999777755555	≥.060	555557777799
9999977777777755555555	≥.055	55555555577799999
999999999777777555	≥.050	99999999 55555555777777777777999
99997777777555	≥.045	77777999
99999977	≥.040	5779999999
9	≥.035	9

Note. The numbers 5, 7, and 9 indicate the three levels of N/p.

the $\dagger$s are above the upper limit of Bradley's (1978) liberal criterion ($.025 \leq \dagger \leq .075$) for Johansen's and James' tests respectively; for both tests all the $\dagger$s are within the criterion when $N/p = 7$ or 9.

In Figures 2, 3, and 4, frequency distributions of $\dagger$ for the normal distribution are reported in terms of relationship between sample sizes and v-c matrices for $N/p = 5$, 7, and 9, respectively. In Figure 2, where $N/p = 5$, the $\dagger$s from both the positive and the equal-sample-size conditions are well within Bradley's (1978) criterion for both tests; half of the $\dagger$s from the negative condition are above the criterion for Johansen's test and one third for James' test. In Figures 3 and 4, where $N/p = 7$ and 9 respectively, the $\dagger$s are within the criterion for all three relationships and for both tests. In addition, results in Figures 2 and 3 suggest that when $N/p \leq 7$, the $\dagger$s for James' test are more clustered about the nominal α value than are those for Johansen's test.

Frequency distributions of $\dagger$ are reported, in terms of N/p, in Figure 5 for the beta distribution. The mean and standard deviation of $\dagger$ are .065 and .017 for Johansen's test and .063 and .015 for James' test. Compared with the normal distribution, the beta distribution results in more $\dagger$s that are above the upper limit of the interval (.0403, .0597): 58% of the $\dagger$s for Johansen's test and 54% for James' test. Like the results for the normal distribution, with

Figure 2

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, $N/p = 5$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .05$

Johansen's test		James' test
---	≥.085	-
-	≥.080	-
--	≥.075	--
--	≥.070	
+--	≥.065	---
++-==	≥.060	---++
++++-===	≥.055	===--++++
+++	≥.050	===+++++
++=	≥.045	
	≥.040	+
	≥.035	

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 3

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, $N/p = 7$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .05$

Johansen's test		James' test
	≥.085	
	≥.080	
	≥.075	
-	≥.070	
-	≥.065	
---=	≥.060	==---
+++------=	≥.055	--+
+++--=	≥.050	==-------++++++
+++++==	≥.045	=++++
+=	≥.040	=+
	≥.035	

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 4

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, $N/p = 9$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .05$

Johansen's test		James' test
	≥.085	
	≥.080	
	≥.075	
	≥.070	
--	≥.065	-
---	≥.060	--
++--=	≥.055	=--++
++++---==	≥.050	==----+++++
+++=	≥.045	=-+
+++--=	≥.040	=--++++
=	≥.035	=

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 5

Frequency Distributions of Estimated Type I Error Rates for the Beta Distribution, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .05$

Johansen's test		James' test
97775555555	≥.085	555557
55	≥.080	59
755	≥.075	55577
999997777775555	≥.070	55555777799
977555555	≥.065	555557777999
999999977555	≥.060	5557799999999
999997777555	≥.055	55557777999
99999977775	≥.050	557777779999999
99977777755	≥.045	5577779999
9977	≥.040	7799
	≥.035	7

Note. The numbers 5, 7, and 9 indicate the three levels of N/p.

the beta distribution there are more $\dagger$s above the upper limit of the interval for N/p = 5 than for N/p = 7 or N/p = 9. By Bradley's (1978) criterion, when α = .05 (.025 $\leq \dagger \leq$ .075), 37%, 10%, and 3% of the $\dagger$s for Johansen's test and 30%, 10%, and 3% of the $\dagger$s for James' test are above the criterion when N/p = 5, 7, and 9, respectively. For both the normal and beta distributions, a review of Figures 1 and 5 indicates that $\dagger$ tends to be closer to .05 as N/p increases. However, the improvement is not as good for the beta distribution as it is for the normal distribution.

In Figures 6, 7, and 8, frequency distributions of $\dagger$ for the beta distribution are reported in terms of relationship between sample sizes and v-c matrices for N/p = 5, 7, and 9, respectively. In Figure 6, where N/p = 5, 100%, 67%, and 25% of the $\dagger$s from the positive condition, the equal-sample-size condition, and the negative condition, respectively, are within Bradley's (1978) criterion for Johansen's test and 100%, 83%, and 33% are within Bradley's criterion for James' test. In Figure 7, where N/p = 7, all of the $\dagger$s from both positive and equal-sample-size conditions and 75% of the $\dagger$s from negative condition fall within the criterion for both tests. In Figure 8, where N/p = 9, all $\dagger$s are within the criterion except for one $\dagger$ from the negative condition for both tests. A review of Figures 2, 3, and 4 for the normal distribution and Figures 6, 7, and 8 for the beta distribution indicates that the

Figure 6

Frequency Distributions of Estimated Type I Error Rates for the Beta Distribution, $N/p = 5$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .05$

Johansen's test		James' test
-------	≥.085	-----
--	≥.080	=
==	≥.075	---
++-=	≥.070	==--+
+++++-	≥.065	-++++
++-	≥.060	+++
+==	≥.055	=-++
+	≥.050	=+
+=	≥.045	=+
	≥.040	
	≥.035	

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 7

Frequency Distributions of Estimated Type I Error Rates for the Beta Distribution, N/p = 7, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .05$

Johansen's test		James' test
---	≥.085	-
	≥.080	
-	≥.075	--
+-----	≥.070	=--+
-=	≥.065	----
+=	≥.060	=-
+++=	≥.055	-+++
+--=	≥.050	===-++
++++==	≥.045	=+++
++	≥.040	++
	≥.035	+

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 8

Frequency Distributions of Estimated Type I Error Rates for the Beta Distribution, $N/p = 9$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .05$

Johansen's test		James' test
-	≥.085	
	≥.080	-
	≥.075	
-----	≥.070	--
-	≥.065	---
++++--=	≥.060	=---++++
+++--	≥.055	-++
+-====	≥.050	===--++
+++	≥.045	=+++
+=	≥.040	=+
	≥.035	

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

performance of both tests is worst when N/p = 5 and the relationship is negative. However, as N/p increases, the performance of each test improves--more so for the normal distribution than for the beta distribution.

Frequency distributions of $\dagger$ under the negative condition are reported in Figure 9 for the exponential distribution when N/p = 9. The mean and standard deviation of $\dagger$ are .085 and .02 for Johansen's test and .082 and .0184 for James' test. When compared with Figure 4 for the normal distribution and Figure 8 for the beta distribution, Figure 9 shows that the exponential distribution results in many more $\dagger$s above the upper limit of the interval (.0403, .0597): 92% of the $\dagger$s for both tests. By Bradley's (1978) criterion ($.025 \leq \dagger \leq .075$), 30% of the observations are within the criterion for Johansen's test and 42% for James' test. The lowest and highest values are .047 and .11 for Johansen's test and .047 and .107 for James' test.

Frequency distributions of $\dagger$ under the negative condition are reported, in terms of N/p, in Figure 10 for the normal distribution and $\Sigma_2 = d^2\Sigma_1$. The mean and standard deviation of $\dagger$ are .078 and .028 for Johansen's test and .071 and .024 for James' test. Further, 72% of the $\dagger$s are above the interval (.0403, .0597) for Johansen's test and 56% for James' test. By Bradley's (1978) criterion ($.025 \leq \dagger \leq .075$), 58% of the $\dagger$s are within the criterion for Johansen's test and 69% for James' test. In particular, for

Figure 9

Frequency Distributions of Estimated Type I Error Rates for the Exponential Distribution, $N/p = 9$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .05$

Johansen's test		James' test
-	≥.110	
--	≥.105	-
	≥.100	--
	≥.095	
---	≥.090	---
-	≥.085	
	≥.080	
-	≥.075	--
-	≥.070	-
-	≥.065	-
-	≥.060	-
	≥.055	
	≥.050	
-	≥.045	-

Note. All results are for the negative relationship between sample sizes and v-c matrices.

Figure 10

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, Negative Condition, $\Sigma_2 = d^2\Sigma_1$, and $\alpha = .05$

Johansen's test		James' test
555	≥.110	555
5	≥.105	
75	≥.100	
5	≥.095	
	≥.090	
75	≥.085	5555
9555	≥.080	577
977	≥.075	7
7	≥.070	59
977	≥.065	579
999775	≥.060	7799
999775	≥.055	557777999
9997	≥.050	799999
	≥.045	7

Note. The numbers 5, 7, and 9 indicate the three levels of N/p.

both tests, most of the ts are above the criterion when N/p = 5. On the other hand, the majority of the ts are within the criterion for both tests when N/p = 7 and 9.

Corresponding to Figure 10, Figure 11 shows the frequency distributions of t for the normal distribution when $\Sigma_2 \neq d^2\Sigma_1$. The mean and standard deviation of ts are .064 and .012 for Johansen's test and .062 and .014 for James' test. Further, for Johansen's test 58% of the ts are above the upper limit of the interval (.0403, .0597) and for James' test 44% are above the upper limit. By Bradley's (1978) criterion, 14% of the ts are above the criterion for Johansen's test and 11% for James' test. In particular, for both tests, N/p = 5 accounts for all of the ts that are above the criterion. A comparison between Figure 10 for $\Sigma_2 = d^2\Sigma_1$ and Figure 11 for $\Sigma_2 \neq d^2\Sigma_1$ indicates that type of heterogeneity of v-c matrices does have an impact on t. Specifically, for each of the three values of N/p, t is substantially larger when $\Sigma_2 = d^2\Sigma_1$ than when $\Sigma_2 \neq d^2\Sigma_1$.

Summary. In sum, when $\alpha = .05$, James' test performs, to some degree, better than Johansen's test. (This same pattern is observed with $\alpha = .10$, see Appendix B. With $\alpha = .01$, Johansen's test performs very slightly better than James' test, see Appendix A). Apart from this distinction, the major patterns of Type I error rates are similar for all three levels of significance. Hence, the results will

Figure 11

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, Negative Condition, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .05$

Johansen's test		James' test
	≥.110	
	≥.105	
	≥.100	
	≥.095	
	≥.090	
555	≥.085	5
5	≥.080	5
55	≥.075	55
755	≥.070	
99755	≥.065	5559
9997775	≥.060	55577799
99777775	≥.055	557799
99977	≥.050	77777779999
	≥.045	9
99	≥.040	99

Note. The numbers 5, 7, and 9 indicate the three levels of N/p.

be summarized as a whole, regardless of the level of significance.

With either the normal distribution or the beta distribution when $\Sigma_2 \neq d^2\Sigma_1$, the performance of both tests generally improves as N/p increases, with their error rates being much larger when N/p = 5 than when N/p = 7 or 9. In particular, when N/p = 5, the negative condition tends to produce extremely liberal results. Furthermore, the results under the negative condition are more liberal than those under either the positive condition or the equal-sample-size condition, irrespective of the value of N/p. On the other hand, there appears no obvious difference in error rates between the positive condition and the equal-sample-size condition. By comparison, the normal distribution tends to produce smaller error rates than those from the beta distribution.

With the exponential distribution when $\Sigma_2 \neq d^2\Sigma_1$ and N/p = 9, under the negative condition both tests tend to produce much larger error rates than those for the normal and beta distributions. Since the skewness and kurtosis of both the normal and beta distributions are smaller than those of the exponential distribution, this confirms the findings of Algina and Oshima (1990b) that distributions with large skewness and kurtosis tend to produce large error rates.

Similar to the results for the normal distribution and $\Sigma_2 \neq d^2\Sigma_1$, with the normal distribution and $\Sigma_2 = d^2\Sigma_1$ Type I

error rates decline as N/p increases. However, for each value of N/p, $\uparrow$ is substantially larger when $\Sigma_2 = d^2\Sigma_1$ than when $\Sigma_2 \neq d^2\Sigma_1$. Thus, type of heterogeneity of v-c matrices does appear to have an impact on the error rates. This finding is in contradiction with that of Tang (1989), who found that type of heterogeneity of v-c matrices had little effect on the error rates for both tests.

Significance Tests

As shown in Table 4, the design of the study was not complete with respect to the heteroscedasticity pattern for v-c matrices, distribution, N/p, and the relationship between sample size and v-c matrices. To accomodate the incomplete design, three analyses were conducted. In the first, $\uparrow$ for $\Sigma_2 \neq d^2\Sigma_1$ and the normal and beta distributions was analyzed. In the second, $\uparrow$ for $\Sigma_2 \neq d^2\Sigma_1$, the three distributions, N/p = 9, and the negative relationship was analyzed. In the third, $\uparrow$ for the normal distribution, the negative realtionship, and both $\Sigma_2 \neq d^2\Sigma_1$ and $\Sigma_2 = d^2\Sigma_1$ was analyzed. Each analysis was conducted three times, once using $\uparrow$ for $\alpha = .01$, once for $\alpha = .05$, and once for $\alpha = .10$. In general the results for $\alpha = .01$, .05, and .10 were similar, and only the results for $\alpha = .05$ are reported in this chapter. Summary ANOVA tables for $\alpha = .01$ and $\alpha = .10$ are presented in Appendix C. Because of memory limitations in SAS and because of the unlikely interpretability of

Table 4

Factorial Combinations of Distribution, N/p, V-C Heterogeneity, and Relationship between Sample Sizes and V-C Matrices

Heterogeneity Pattern	Distribution	N/p	Relationship		
			$n_1 < n_2$[a]	$n_1 = n_2$	$n_1 > n_2$[b]
$\Sigma_2 \neq d^2\Sigma_1$	Normal	5,7,9	*	*	*
	Beta	5,7,9	*	*	*
	Exponential	9			*
$\Sigma_2 = d^2\Sigma_1$	Normal	5,7,9			*
	Beta				
	Exponential				

Note. A * indicates a factorial combination used in the study.

[a] positive relationship between sample sizes and v-c matrices.

[b] negative relationship between sample sizes and v-c matrices.

three-way and higher order interactions, only main effects and two-way interactions were tested.

Analysis 1. The F values and p values for the main effects and two-way interactions are given in Table 5. As the table shows, each of the seven factors--test (T), distribution (D), relationship (R), number of variables (p), ratio of N to p (N/p), sample-size ratio (n_1:n_2), and degree of heterogeneity of v-c matrices (d)--had a significant interaction with at least two other factors. The value of R-square for the model was .8464. This implies that about 85% of the variability in the data is explained by the main and interaction effects, and little is lost by pooling the higher-order interactions with the error term.

Least square means for the T main effect and the T x R and T x p interactions are given in Table 6. With $F_{1,308} = 6.83$ and $p < .05$ for T, it is evident that James' test performs significantly better than Johansen's test.

The R factor interacts significantly with T but does not change the direction of the main effect of T. That is, James' test performs better than Johansen's test, regardless of the level of R. In particular, the superiority of James' test over that of Johansen's test is most obvious when the relationship is negative. The p factor also interacts significantly with T, but changes the direction of the main effect of T. When p is either 6 or 10, James' test performs better than Johansen's test. However, when p is 2,

Table 5

Summary ANOVA Table: Analysis 1 and $\alpha = .05$

Source of Variation	Degrees of Freedom	F Value	PR > F
Test (T)	1	6.83	.0094*
Distribution (D)	1	136.99	.0001*
Relationship (R)	1	432.96	.0001*
p	2	1.18	.3081
N/p	2	134.63	.0001*
$n_1:n_2$	1	131.05	.0001*
d	1	122.36	.0001*
T x D	1	0.00	.9466
T x R	1	7.33	.0072*
T x p	2	9.75	.0001*
T x N/p	2	1.90	.1507
T x $n_1:n_2$	1	2.19	.1403
T x d	1	0.77	.3795
D x R	1	38.36	.0001*
D x p	2	3.36	.0361*
D x N/p	2	4.28	.0147*
D x $n_1:n_2$	1	10.41	.0014*
D x d	1	54.80	.0001*
R x p	2	4.02	.0190*
R x N/p	2	29.91	.0001*
R x $n_1:n_2$	1	29.10	.0001*
R x d	1	39.23	.0001*
p x N/p	4	1.20	.3104
p x $n_1:n_2$	2	2.87	.0583
p x d	2	10.02	.0001*
N/p x $n_1:n_2$	2	26.52	.0001*
N/p x d	2	3.92	.0209*
$n_1:n_2$ x d	1	0.03	.8574
Error	308		

Note. When testing the two factors, R and $n_1:n_2$, or any interaction involving either of these two factors, the data under equal-sample-size condition are not taken into consideration by GLM in SAS. Thus, the degrees of freedom have to be reduced by one or adjusted accordingly.

* $p < .05$.

Table 6

Main Effect Means and First-Order Interaction Means for the Test Factor: Analysis 1

Main Effect		Johansen	James
Test (T)		.0611	.0591
Interaction	**Levels**	**Johansen**	**James**
T x R	$n_1 < n_2$[a]	.0544	.0540
	$n_1 = n_2$	.0551	.0547
	$n_1 > n_2$[b]	.0708	.0665
T x p	2	.0595	.0612
	6	.0608	.0581
	10	.0630	.0581

[a] positive relationship between sample sizes and v-c matrices.

[b] negative relationship between sample sizes and v-c matrices.

Johansen's test performs slightly better than James' test. In addition, as p increases there is a slight tendency for the performance of Johansen's test to decline and that of James' test to improve.

Least square means for the D main effect and the D x R, D x p, D x N/p, D x n_1:n_2, and D x d interactions are given in Table 7. With $F_{1,308} = 136.99$ and $p < .05$, the factor D has a significant effect on the error rates. That is, the error rates from the normal distribution are much closer to the nominal α level than are those from the beta distribution.

The factors R, p, N/p, n_1:n_2, and d interact significantly with D. However, none of these factors modify the direction of the main effect of D. In fact, at each level of each of these factors, the mean error rates are smaller when the distribution is normal than when it is beta. In particular, the difference in mean error rates between these two distributions is quite substantial when the relationship is negative, when p = 2, when N/p = 5, when n_1:n_2 = 2:1, or when d = 3. Furthermore, the discrepancy in mean error rates between the normal and beta distributions decreases as p increases, but increases as either n_1:n_2 or d increases.

Least square means for the R main effect and the R x T, R x D, R x p, R x N/p, R x n_1:n_2, and R x d interactions are given in Table 8. With $F_{1,308} = 432.96$ and $p < .05$, it is

Table 7

Main Effect Means and First-Order Interaction Means for the Distribution Factor: Analysis 1

Main Effect		Normal	Beta
Distribution (D)		.0560	.0642
Interaction	**Levels**	**Normal**	**Beta**
D x R	$n_1 < n_2$[a]	.0519	.0565
	$n_1 = n_2$	.0524	.0575
	$n_1 > n_2$[b]	.0620	.0753
D x p	2	.0554	.0654
	6	.0552	.0637
	10	.0575	.0636
D x N/p	5	.0626	.0733
	7	.0539	.0607
	9	.0517	.0586
D x n_1:n_2	1:1	.0524	.0575
	1.5:1	.0541	.0608
	2:1	.0598	.0710
D x d	1.5	.0546	.0582
	3.0	.0575	.0703

[a] positive relationship between sample sizes and v-c matrices.

[b] negative relationship between sample sizes and v-c matrices.

Table 8

Main Effect Means and First-Order Interaction Means for the Relationship between Sample Sizes and V-C Matrices Factor: Analysis 1

Main Effect	$n_1 < n_2$[a]	$n_1 = n_2$	$n_1 > n_2$[b]
Relationship (R)	.0542	.0549	.0687

Interaction	Levels	$n_1 < n_2$	$n_1 = n_2$	$n_1 > n_2$
R x T	Johansen	.0544	.0551	.0708
	James	.0540	.0547	.0665
R x D	Normal	.0519	.0524	.0620
	Beta	.0565	.0575	.0753
R x p	2	.0539	.0519	.0711
	6	.0536	.0559	.0671
	10	.0551	.0570	.0678
R x N/p	5	.0589	.0607	.0805
	7	.0514	.0541	.0647
	9	.0522	.0499	.0608
R x n_1:n_2	1:1		.0549	
	1.5:1	.0521		.0628
	2:1	.0563		.0745
R x d	1.5	.0525	.0519	.0626
	3.0	.0559	.0580	.0748

[a] positive relationship between sample sizes and v-c matrices.

[b] negative relationship between sample sizes and v-c matrices.

evident that R has a significant impact on the error rates. Specifically, the error rates are substantially larger under the negative condition than under the equal-sample-size or the positive condition. Further, the error rates under the positive condition are slightly smaller than those under the equal-sample-size condition.

The factors T, D, p, N/p, n_1:n_2, and d interact significantly with R. However, none of these factors change the direction of the main effect of R in any important way. That is, the mean error rate under the negative condition is larger than those under either the equal-sample-size or the positive condition. In addition, the mean error rates under the positive condition remain slightly smaller than those under the equal-sample-size condition except when p = 2, when d = 1.5, or when N/p = 9.

Least square means for the p main effect and the p x T, p x D, p x R, and p x d interactions are given in Table 9. With $F_{2,308} = 1.18$ and $p > .05$, it is evident that p does not have a significant effect on the error rates. That is, there is little difference in error rates among the three levels of p.

The factors T, D, R, and d interact significantly with p. The pattern of mean error rates for p is erratic over the levels of these factors: As the value of p changes, the mean error rates increase in some situations but decrease in others. For example, as p increases from 2 to 6 to 10, the

Table 9

Main Effect Means and First-Order Interaction Means for the p Factor: Analysis 1

Main Effect		2	6	10
p		.0604	.0594	.0606
Interaction	**Levels**	**2**	**6**	**10**
p x T	Johansen	.0595	.0608	.0630
	James	.0612	.0581	.0581
p x D	Normal	.0554	.0552	.0575
	Beta	.0654	.0637	.0636
p x R	$n_1 < n_2$[a]	.0539	.0536	.0551
	$n_1 = n_2$	.0519	.0559	.0570
	$n_1 > n_2$[b]	.0711	.0671	.0678
p x d	1.5	.0548	.0559	.0584
	3.0	.0659	.0630	.0627

[a] positive relationship between sample sizes and v-c matrices.

[b] negative relationship between sample sizes and v-c matrices.

mean error rate for Johansen's test increases from .0595 to .0608 to .0630 correspondingly, whereas that for James' test decreases from .0612 to .0581. Similarly, as p increases from 2 to 6 to 10, the mean error rate under the equal-sample-size condition increases from .0519 to .0559 to .0570, whereas that under the negative condition the mean error rate changes from .0711 to .0671 and to .0678. Further, an interesting phenomenon is that as p increases, the mean error rate increases when d = 1.5 and decreases when d = 3. Hence, the discrepancy in mean error rates between d = 1.5 and d = 3 decreases as p increases.

Least square means for the N/p main effect and the N/p x D, N/p x R, N/p x n_1:n_2, and N/p x d interactions are given in Table 10. With $F_{2,308} = 134.63$ and $p < .05$, it is evident that N/p has a significant effect on the error rates. In particular, as N/p increases, the error rates decrease--with the decline being much larger between N/p = 5 and N/p = 7 than between N/p = 7 and N/p = 9.

The factors D, R, n_1:n_2, and d interact significantly with N/p. However, in most cases, these factors do not change the pattern of the main effect of N/p. That is, the mean error rates decrease as N/p increases, with the discrepancy in mean error rates being much larger between N/p = 5 and N/p = 7 than between N/p = 7 and N/p = 9. However, there are two exceptions to this pattern, namely, when the relationship is positive the mean error rate

Table 10

Main Effect Means and First-Order Interaction Means for the N/p Factor: Analysis 1

Main Effect		5	7	9
N/p		.0679	.0573	.0552
Interaction	**Levels**	**5**	**7**	**9**
N/p x D	Normal	.0626	.0539	.0517
	Beta	.0733	.0607	.0586
N/p x R	$n_1 < n_2$[a]	.0589	.0514	.0522
	$n_1 = n_2$	.0607	.0541	.0499
	$n_1 > n_2$[b]	.0805	.0647	.0608
N/p x n_1:n_2	1:1	.0607	.0541	.0499
	1.5:1	.0624	.0546	.0553
	2:1	.0770	.0616	.0576
N/p x d	1.5	.0630	.0545	.0516
	3.0	.0728	.0601	.0587

[a] positive relationship between sample sizes and v-c matrices.

[b] negative relationship between sample sizes and v-c matrices.

increases slightly as N/p increases from 7 to 9, and when $n_1:n_2$ = 1.5:1 the mean error rate increases slightly as N/p increases from 7 to 9.

Least square means for the $n_1:n_2$ main effect and the $n_1:n_2$ x D, $n_1:n_2$ x R, $n_1:n_2$ x N/p interactions are given in Table 11. With $F_{1,308}$ = 131.05 and $p < .05$, the factor $n_1:n_2$ has a significant impact on the error rates. Specifically, as the sample size ratio increases, the error rates increase--with the increase being larger when $n_1:n_2$ increases from 1.5:1 to 2:1 than when it increases from 1:1 to 1.5:1.

The factors D, R, and N/p interact significantly with $n_1:n_2$. However, in most cases, these factors do not change the pattern of the main effect of $n_1:n_2$. That is, the increase in mean error rates is larger when $n_1:n_2$ increases from 1.5:1 to 2:1 than when it increases from 1:1 to 1.5:1. Nevertheless, there is one exception, when N/p = 9, the increase in mean error rates is slightly larger when $n_1:n_2$ increases from 1:1 to 1.5:1 than when it increases from 1.5:1 to 2:1.

Least square means for the d main effect and the d x D, d x R, d x p, and d x N/p interactions are given in Table 12. With $F_{1,308}$ = 122.36 and $p < .05$, the error rates are significantly larger when d = 3 than when d = 1.5. The factors D, R, p, and N/p interact significantly with d. However, none of these factors modify the direction of the

Table 11

Main Effect Means and First-Order Interaction Means for the n_1:n_2 Factor: Analysis 1

Main Effect		1:1	1.5:1	2:1
n_1:n_2		.0549	.0574	.0654

Interaction	Levels	1:1	1.5:1	2:1
n_1:n_2 x D	Normal	.0524	.0541	.0598
	Beta	.0575	.0608	.0710
n_1:n_2 x R	$n_1 < n_2$[a]		.0521	.0563
	$n_1 = n_2$	.0549		
	$n_1 > n_2$[b]		.0628	.0745
n_1:n_2 x N/p	5	.0607	.0624	.0770
	7	.0541	.0546	.0616
	9	.0499	.0553	.0576

[a] positive relationship between sample sizes and v-c matrices.

[b] negative relationship between sample sizes and v-c matrices.

Table 12

Main Effect Means and First-Order Interaction Means for the d Factor: Analysis 1

Main Effect		1.5	3.0
d		.0564	.0639
Interaction	**Levels**	**1.5**	**3.0**
d x D	Normal	.0546	.0575
	Beta	.0582	.0703
d x R	$n_1 < n_2$[a]	.0525	.0559
	$n_1 = n_2$	.0519	.0580
	$n_1 > n_2$[b]	.0626	.0748
d x p	2	.0548	.0659
	6	.0559	.0630
	10	.0584	.0627
d x N/p	5	.0630	.0728
	7	.0545	.0601
	9	.0516	.0587

[a] positive relationship between sample sizes and v-c matrices.

[b] negative relationship between sample sizes and v-c matrices.

main effect of d. That is, the mean error rates remain relatively larger for d = 3 than for d = 1.5, irrespective of the levels of the interacting factors. In particular, the difference in mean error rates between d = 1.5 and d = 3 is quite substantial when the distribution is beta, when the relationship is negative, when p = 2, or when N/p = 5.

Analysis 2. The F values and p values of the main and two-way interactions are given in Table 13. Because only the conditions in which the relationship is negative and N/p = 9 are investigated, main and interaction effects here involve the factors T, D, p, n_1:n_2, and d. As Table 13 indicates, except for the T and T x p effects, the factors that were significant in the preceding section are also significant in this section.

Least square means for main effects and interactions are given in Tables 14 through 18. With the exception of that for T, the general patterns of the main effects and interaction effects for D, n_1:n_2, d, and p are quite similar to those in the previous section. In other words, the mean error rates are smaller for the normal distribution than for the beta distribution and in turn the mean error rates are smaller for the beta distribution than for the exponential distribution, irrespective of the levels of p, n_1:n_2, and d; the mean error rates are smaller for n_1:n_2 = 1.5:1 than for n_1:n_2 = 2:1, irrespective of the levels of D; the mean error rates are smaller for d = 1.5 than for d = 3.0, irrespective

Table 13

Summary ANOVA Table: Analysis 2 and $\alpha = .05$

Source of Variation	Degrees of Freedom	F Value	PR > F
Test (T)	1	3.00	.0901
Distribution (D)	2	150.01	.0001*
p	2	0.75	.4776
$n_1:n_2$	1	41.92	.0001*
d	1	150.03	.0001*
T x D	2	0.01	.9911
T x p	2	0.71	.4974
T x $n_1:n_2$	1	0.61	.4374
T x d	1	0.33	.5666
D x p	4	3.58	.0128*
D x $n_1:n_2$	2	12.76	.0001*
D x d	2	17.97	.0001*
p x $n_1:n_2$	2	0.78	.4629
p x d	2	12.02	.0001*
$n_1:n_2$ x d	1	0.68	.4138
Error	45		

* $p < .05$.

Table 14

Main Effect Means and First-Order Interaction Means for the Distribution Factor: Analysis 2

Main Effect	Normal	Beta	Exponential
Distribution (D)	.0551	.0664	.0835

Interaction	Levels	Normal	Beta	Exponential
D x p	2	.0554	.0694	.0819
	6	.0550	.0667	.0798
	10	.0550	.0631	.0888
D x n_1:n_2	1.5:1	.0549	.0621	.0749
	2:1	.0554	.0706	.0921
D x d	1.5	.049	.0611	.0695
	3.0	.060	.0716	.0975

Table 15

Main Effect Means and First-Order Interaction Means for the p Factor: Analysis 2

Main Effect		2	6	10
p		.0689	.0672	.0690
Interaction	**Levels**	**2**	**6**	**10**
p x D	Normal	.0554	.0550	.0550
	Beta	.0694	.0667	.0631
	Exponential	.0819	.0798	.0888
p x d	1.5	.0595	.0556	.0652
	3.0	.0783	.0788	.0727

Table 16

Main Effect Means and First-Order Interaction Means for the $n_1:n_2$ Factor: Analysis 2

Main Effect		1.5:1	2:1
$n_1:n_2$		.0640	.0727
Interaction	**Levels**	**1.5:1**	**2:1**
$n_1:n_2$ x D	Normal	.0549	.0554
	Beta	.0621	.0706
	Exponential	.0749	.0921

Table 17

Main Effect Means and First-Order Interaction Means for the d Factor: Analysis 2

Main Effect		1.5	3.0
d		.0601	.0766
Interaction	**Levels**	**1.5**	**3.0**
d x D	Normal	.0496	.0607
	Beta	.0611	.0716
	Exponential	.0695	.0975
d x p	2	.0595	.0783
	6	.0556	.0788
	10	.0652	.0727

Table 18

Main Effect Means for the Test Factor: Analysis 2

Main Effect	Johansen	James
Test (T)	.0695	.0672

of the levels of D and p; and lastly the mean error rates are not much different among the three levels of p, with the pattern of the mean error rates getting erratic when p interacts with D or d. On the other hand, for the factor T, James' test performs slightly, though not significantly, better than Johansen's test.

Analysis 3. The F values and p values of the main effects and two-way interactions are given in Table 19. Because only the conditions in which the relationship between sample sizes and v-c matrices is negative are investigated, main and interaction effects involve the factors T, p, N/p, n_1:n_2, d, and type of heterogeneity of v-c matrices (TH). Excluding the TH factor from consideration, the acceptance/rejection decisions for the main and interaction effects--except for those about p x d and n_1:n_2 x d interactions--are the same as those in the first section.

Least square means for the T, p, N/p, n_1:n_2, and d main effects and their two-way interactions are given in Tables 20 through 24. In general, the patterns of the main and interaction effects of these factors are quite similar to those in the first section. That is, the mean error rate is smaller for Johansen's test than for James' test when p = 2 and larger when p = 6 or 10; the mean error rates are smaller for large N/p than for small N/p, irrespective of the levels of n_1:n_2, d, and TH; the mean error rates are

Table 19

Summary ANOVA Table: Analysis 3 and $\alpha = .05$

Source of Variation	Degrees of Freedom	F Value	PR > F
Test (T)	1	14.76	.0002*
p	2	1.03	.3601
N/p	2	130.81	.0001*
$n_1:n_2$	1	79.95	.0001*
d	1	91.69	.0001*
Type of heterogeneity of v-c matrices (TH)	1	72.84	.0001*
T x p	2	6.97	.0014*
T x N/p	2	1.10	.3353
T x $n_1:n_2$	1	1.49	.2249
T x d	1	1.08	.3003
T x TH	1	0.97	.3273
p x N/p	4	0.75	.5577
p x $n_1:n_2$	2	0.89	.4126
p x d	2	0.93	.3970
p x TH	2	2.39	.0964
N/p x $n_1:n_2$	2	24.65	.0001*
N/p x d	2	9.88	.0001*
N/p x TH	2	15.69	.0001*
$n_1:n_2$ x d	1	7.90	.0059*
$n_1:n_2$ x TH	1	17.41	.0001*
d x TH	1	21.67	.0001*
Error	109		

* $p < .05$.

Table 20

Main Effect Means and First-Order Interaction Means for the Test Factor: Analysis 3

Main Effect	Johansen	James
Test (T)	.0711	.0655

Interaction	Levels	Johansen	James
T x p	2	.0667	.0683
	6	.0710	.0642
	10	.0756	.0639

Table 21

Main Effect Means and First-Order Interaction Means for the p Factor: Analysis 3

Main Effect		2	6	10
p		.0675	.0676	.0698
Interaction	**Levels**	**2**	**6**	**10**
p x T	Johansen	.0667	.0710	.0756
	James	.0683	.0642	.0639

Table 22

Main Effect Means and First-Order Interaction Means for the N/p Factor: Analysis 3

Main Effect		5	7	9
N/p		.0848	.0626	.0575

Interaction	Levels	5	7	9
N/p x n_1:n_2	1.5:1	.0710	.0591	.0551
	2:1	.0986	.0660	.0599
N/p x d	1.5	.0733	.0568	.0536
	3.0	.0963	.0683	.0613
N/p x TH	$\Sigma_2 \neq d^2\Sigma_1$	.0729	.0581	.0551
	$\Sigma_2 = d^2\Sigma_1$	.0968	.0670	.0599

Table 23

Main Effect Means and First-Order Interaction Means for the n_1:n_2 Factor: Analysis 3

Main Effect		1.5:1	2:1
n_1:n_2		.0617	.0749
Interaction	**Levels**	**1.5:1**	**2:1**
n_1:n_2 x N/p	5	.0710	.0986
	7	.0591	.0660
	9	.0551	.0599
n_1:n_2 x d	1.5	.0568	.0658
	3.0	.0667	.0839
n_1:n_2 x TH	$\Sigma_2 \neq d^2\Sigma_1$	.0585	.0655
	$\Sigma_2 = d^2\Sigma_1$	.0649	.0842

Table 24

Main Effect Means and First-Order Interaction Means for the d Factor: Analysis 3

Main Effect	1.5	3.0
d	.0613	.0753

Interaction	Levels	1.5	3.0
d x N/p	5	.0733	.0963
	7	.0568	.0683
	9	.0536	.0613
d x $n_1:n_2$	1.5:1	.0568	.0667
	2:1	.0658	.0839
d x TH	$\Sigma_2 \neq d^2\Sigma_1$	.0584	.0656
	$\Sigma_2 = d^2\Sigma_1$	.0641	.0850

smaller for $n_1:n_2$ = 1.5:1 than for $n_1:n_2$ = 2:1, irrespective of the levels of N/p, d, and TH; the mean error rates are smaller for d = 1.5 than for d = 3.0, irrespective of the levels of N/p, $n_1:n_2$, and TH; and lastly the mean error rates are not much different among the three levels of p, with a tendency for the performance of Johansen's test to decline and that of James' test to improve as p increases.

Least square means for the factor and for the TH x N/p, TH x $n_1:n_2$, and TH x d interactions are given in Table 25. With $F_{1,109} = 72.84$ and $p < .05$, the error rates are significantly different between the two levels of TH. The factors N/p, $n_1:n_2$, and d interact significantly with TH but do not change the direction of the main effect of TH. That is, the mean error rates are larger for $\Sigma_2 = d^2\Sigma_1$ than for $\Sigma_2 \neq d^2\Sigma_1$, irrespective of the levels of N/p, d, and $n_1:n_2$.

For N/p, $n_1:n_2$, d, and TH, none of the interacting factors change the direction of the main effect. However, the magnitude of the difference in mean error rates between any two levels of a factor depends on the levels of the interacting factors. Specifically, the difference in mean error rates between any two levels of N/p increases when $n_1:n_2$ increases from 1.5:1 to 2:1, when d increases from 1.5 to 3.0, or when TH changes from $\Sigma_2 \neq d^2\Sigma_1$ to $\Sigma_2 = d^2\Sigma_1$; the difference between the two levels of $n_1:n_2$ increases when N/p decreases from 9 to 7 to 5, when d increases from 1.5 to 3.0, or when TH changes from $\Sigma_2 \neq d^2\Sigma_1$ to $\Sigma_2 = d^2\Sigma_1$; the

Table 25

Main Effect Means and First-Order Interaction Means for the Type of Heterogeneity of V-C Matrices Factor: Analysis 3

Main Effect	$\Sigma_2 \neq d^2\Sigma_1$	$\Sigma_2 = d^2\Sigma_1$
Type of Heterogeneity of V-C Matrices (TH)	.0620	.0746

Interaction	Levels	$\Sigma_2 \neq d^2\Sigma_1$	$\Sigma_2 = d^2\Sigma_1$
TH x N/p	5	.0729	.0968
	7	.0581	.0670
	9	.0551	.0599
TH x n_1:n_2	1.5:1	.0585	.0649
	2:1	.0655	.0842
TH x d	1.5	.0584	.0641
	3.0	.0656	.0850

difference between the two levels of d increases when N/p decreases from 9 to 7 to 5, when $n_1:n_2$ increases from 1.5:1 to 2:1, or when TH changes from $\Sigma_2 \neq d^2\Sigma_1$ to $\Sigma_2 = d^2\Sigma_1$; and lastly the difference between the two levels of TH increases when N/p decreases from 9 to 7 to 5, when $n_1:n_2$ increases from 1.5:1 to 2:1, or when d increases from 1.5 to 3.0. In contrast, this finding is not so obvious in the cases of the first and second analyses.

Summary. In sum, all the eight factors--ratio of N to p, sample-size ratio, degree of heterogeneity of v-c matrices, type of distribution, type of relationship, type of heterogeneity of v-c matrices, number of variables, and type of test--affect the outcomes of the error rates. Typically, interactions are such that the magnitude of an effect changes over levels of a second factor. In only a few interactions, did the direction of the effect change as a function of the second factor.

The factor N/p has a considerable effect on the error rates. In general, as N/p increases, the error rates decrease. Furthermore, for each level of another factor, the difference in mean error rates between N/p = 5 and N/p = 7 is much larger than that between N/p = 7 and N/p = 9.

On the other hand, the error rates increase as $n_1:n_2$ increases, with the increase being larger when $n_1:n_2$ increases from 1.5:1 to 2:1 than when it increases from 1:1 to 1.5:1. Likewise, the error rates increase as d

increases, no matter what the levels of the other factors are.

The normal distribution results in error rates significantly closer to α than does the beta distribution, regardless of the levels of other factors. In turn, the beta distribution has significantly better Type I error rates than does the exponential distribution, regardless of the levels of other factors.

The error rates are significantly larger under the negative condition than under either the equal-sample-size condition or the positive condition, irrespective of the levels of other factors. Between the equal-sample-size condition and the positive condition, one produces relatively smaller error rates in some situations and larger rates in other situations. However, on the whole, the positive condition seems to have slightly superior Type I error rates relative to the equal-sample-size condition.

The factor TH has a substantial impact on the error rates. Specifically, the error rates are smaller when $\Sigma_2 \neq d^2\Sigma_1$ than when $\Sigma_2 = d^2\Sigma_1$, regardless of the levels of N/p, $n_1{:}n_2$, and d. The effect of TH is most obvious when N/p = 5, when $n_1{:}n_2 = 2{:}1$, or when d = 3.0. This finding is in contradiction with that of Tang (1989) who claimed that TH has little effect on the performance of either James' or Johansen's test.

Consistent with results reported by Tang (1989), the factor p does not seem to have a strong impact on the error rates. However, when other factors are involved, the effect of p is erratic: as p increases, the error rates increase in some situations but decrease in others. On the other hand, contrary to Tang's finding that small value of p leads to small error rates for James' test, the finding here indicates that James' test performs better for large value of p. In contrast, Johansen's test performs better for small values of p.

On the whole, James' test performs slightly better than Johansen's test, with both tests performing better under positive and equal-sample-size conditions than under negative condition. In contrast, Tang (1989) reported that James' test performs better under the negative condition than under the positive condition. However, Tang studied the case of three groups and included a wider variety of heteroscedasticity patterns. Furthermore, both tests perform similarly under the positive and the equal-sample-size conditions, but James' test performs better than Johansen's test under the negative condition. As p increases, the performance of James' test improves and that of Johansen's test declines.

CHAPTER 5
CONCLUSIONS AND RECOMMENDATIONS

Two hundred and twenty-eight simulated experiments--192 with unequal sample sizes and 36 with equal sample sizes--were conducted in this study. The design is complete for the normal and beta distributions when $\Sigma_2 \neq d^2\Sigma_1$ but incomplete for the exponential distribution when $\Sigma_2 \neq d^2\Sigma_1$ and for the normal distribution when $\Sigma_2 = d^2\Sigma_1$.

The results of this study are intended to provide researchers with some guidelines for comparing the means of two populations under variance-covariance heteroscedasticity when the ratio of N to p is less than 10. However, the generalizability of the results is limited by the range of values for the factors.

The Type I error rates for James' test tend to be smaller than those for Johansen's test, with larger differences when the relationship between sample sizes and v-c matrices is negative. An exception to the generalization occurs for p = 2. In this exceptional condition, the mean Type I error rates for James' test is larger than that for Johansen's test.

The Type I error rates improve for both tests as N/p increases from 5 to 9, as relationship between sample sizes

and v-c matrices changes from negative to positive, as the sample-size ratio decreases from 2:1 to 1:1, when sampling from symmetric distributions rather than asymmetric distributions, and when there are variance differences on a subset of the variables in contrast to all of the variables. In addition, some of these factors have an interactive effect on the Type I error rates in some conditions but not in others.

When $\alpha = .05$, Bradley's (1978) liberal criterion for robustness is a Type I error rate in the interval (.025, .075). For the normal and beta distributions, $\alpha = .05$, and $\Sigma_2 \neq d^2\Sigma_1$, Table 26 contains the percentage of estimated Type I error rates within Bradley's interval. For the normal distribution, $\alpha = .05$, and $\Sigma_2 = d^2\Sigma_1$, and for the exponential distribution, $\alpha = .05$, and $\Sigma_2 \neq d^2\Sigma_1$, Table 27 contains the percentage of estimated Type I error rates within Bradley's interval. The results in Table 27 are for the negative condition. In the following conditions, both James' and Johansen's tests had all estimated Type I error rates in this interval: multivariate normal distribution, variance differences on half the variables, and $N/p \geq 7$. For $N/p = 5$, Type I error rates were outside Bradley's interval only for the negative relationship between sample sizes and v-c matrices, and then only for $n_1:n_2 = 2:1$. With data sampled from a multivariate normal distribution and variance differences on all the variables, 100% of the estimated Type

Table 26

Percentage of Estimated Type I Error Rates within Bradley's Criterion for $\alpha = .05$, the Normal and Beta Distributions, and $\Sigma_2 \neq d^2\Sigma_1$

		Normal		Beta	
N/p	n_1 & n_2	Johansen	James	Johansen	James
5	$n_1 = n_2$	100%*	100%*	67%	83%
	$n_1 > n_2$[a]	50%	67%	25%	33%
	$n_1 < n_2$[b]	100%*	100%*	100%*	100%*
7	$n_1 = n_2$	100%*	100%*	100%*	100%*
	$n_1 > n_2$	100%*	100%*	63%	75%
	$n_1 < n_2$	100%*	100%*	100%*	100%*
9	$n_1 = n_2$	100%*	100%*	100%*	100%*
	$n_1 > n_2$	100%*	100%*	92%*	92%*
	$n_1 < n_2$	100%*	100%*	100%*	100%*

Note. A * indicates that the test is relatively robust if, by rule of thumb, at least 90% of estimated Type I error rates are within Bradley's interval.

[a] negative relationship between sample sizes and v-c matrices.

[b] positive relationship between sample sizes and v-c matrices.

Table 27

Percentage of Estimated Type I Error Rates within Bradley's Criterion for the Exponential Distribution When $\Sigma_2 \neq d^2\Sigma_1$ and for the Normal Distribution When $\Sigma_2 = d^2\Sigma_1$: $\alpha = .05$ and the Negative Condition

	Exponential		Normal	
N/p	Johansen	James	Johansen	James
5			17%	33%
7			75%	75%
9	33%	42%	83%	100%*

Note. A * indicates that the test is relatively robust if, by rule of thumb, at least 90% of estimated Type I error rates are within Bradley's interval.

I error rates for James' test were in the interval when N/p = 9. The corresponding percentage for Johansen's test was 83%. As N/p declined to 7 and 5, the percentage of Type I error rates in the interval declined to 75 and 33 for James' test and to 75 and 17 for Johansen's test. For N/p = 7, all Type I error rates outside Bradley's interval occurred for $n_1:n_2 = 2:1$. This was true for both Johansen's and James' tests. For N/p = 5, both Johansen's and James' tests had Type I error rates outside Bradley's interval when $n_1:n_2 \geq 1.5$. Because sample v-c matrices may be an unreliable guide to the type of heterogeneity and the relationship to sample sizes, these results indicate that when either test is used with symmetric multivariate distributions and $1.5 < n_1:n_2 \leq 2.0$, N/p should be at least 9. When $n_1:n_2 \leq 1.5$, N/p can be reduced to seven. Further research with $\Sigma_2 = d^2\Sigma_1$ is needed to establish whether N/p = 5 is sufficient when $n_1 = n_2$. In addition, additional research with other patterns of heteroscedasticity is needed to establish the generality of these guidelines.

When sampling from the beta distributions, the condition in which $\Sigma_2 = d^2\Sigma_1$ was not studied. For conditions in which there were variance differences on half the variables and N/p = 9, James' test resulted in only one estimated Type I error rate outside Bradley's (1978) interval. The estimate was .0815 and occurred for a condition with $n_1:n_2 = 2:1$ and d = 3. Similarly, Johansen's

test had only one estimated Type I error rate outside the interval. This estimate was .0855 and also occurred for the condition in which $n_1:n_2 = 2:1$ and $d = 3$. For $N/p = 7$, Type I error rates outside Bradley's interval occurred only in the negative condition and then only when $n_1:n_2 = 2:1$. For $N/p = 5$, Type I error rates outside Bradley's interval occurred for $n_1:n_2 \geq 1$. Because the sample v-c matrices may be an unreliable guide to the sample size to v-c matrix relationship, these results suggest that when sampling from moderately asymmetric distributions N/p should be at least 9 if $1.5 < n_1:n_2 \leq 2.0$. When $n_1:n_2 \leq 1.5$, $N/p = 7$ may be sufficient. Additional research with $\Sigma_2 = d^2\Sigma_1$ would indicate whether these guidelines are appropriate when there are variance differences on all variables.

When sampling from exponential distributions, no conditions with $\Sigma_2 = d^2\Sigma_1$ or $N/p < 9$ were investigated. The performance of both tests was poor when $N/p = 9$ and there were variance differences on half the variables. For both tests, more than half the estimated Type I error rates were outside Bradley's (1978) interval, and rates outside the interval occurred for $n_1:n_2 \geq 1.5$. Additional research with larger N/p ratios is required to determine the minimum ratio need for the tests to perform adequately.

The performance of both tests declines when sampling from asymmetric distributions. This raises the issue of how well the distributions used in this study represent

asymmetric distributions found in practice. To my knowledge, no large surveys of skewness and kurtosis of multivariate distributions have been published. Micceri (1989) has conducted a study of the skewness of 440 univariate distributions, including 231 distributions of standardized achievement and ability test scores, 35 distributions of criterion-referenced and mastery test scores, 125 distributions of personality test scores, and 49 distributions of gain scores based on the preceding types of test scores. Micceri used skewness and two other criteria to classify distributions with respect to asymmetry. Cutoff scores were set on each criterion and a distribution was classified by the criterion on which it had the most extreme score. The cutoff scores on skewness were .31, .71, and 2.00. Of the 440 distributions, 30.9% were classified at or above the category associated with the .71 cutscore. The absolute value of the skewness of the beta distribution used in this study is .73, indicating that this distribution is not unreasonably skewed compared to distributions found in practice. Of all distributions, 11.4% were classified at the category defined by the 2.00 cutscore. The exponential distribution has skewness 2.00. Thus, even the skewness of exponential distribution is representative of skewness found in practice. Micceri did not use kurtosis to classify distributions and consequently precise comparison of his distributions to the distributions used in this study is

difficult. However, Micceri found that 18% of the distributions have tailweights below that of a normal distribution and 66.8% of the distributions have tailweights above that of a normal distribution. The majority of the former distributions should have negative kurtosis which is consistent with the beta distribution; the majority of latter distributions should have positive kurtosis which is consistent with the exponential distribution.

Micceri's (1989) results are evidence that the distributions used in this study are not clearly unrepresentative of distributions found in practice. Nevertheless, additional studies of the distributional characteristics of multivariate data are clearly in order. Although tentative, one implication of the study is that, in choosing a sample size, researchers need to consider the likely distributional characteristics of the data in selecting sample sizes. This study is a step in providing guidelines for such selections.

APPENDIX A

FREQUENCY DISTRIBUTIONS OF t WHEN $\alpha = .01$

Figure 12

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .01$

Johansen's test		James' test
5	≥.0300	5
	≥.0275	
5	≥.0250	
	≥.0225	
5	≥.0200	555
55	≥.0175	557
997755	≥.0150	55579
99997777775555555555	≥.0125	9999 555555555555577777777779999
99999999 9997777777777775555555	≥.0100	55555777777777999999999
999999997777777555555	≥.0075	55577777799999999
9999977	≥.0050	779999

Note. The numbers 5, 7, and 9 indicate the three levels of N/p.

Figure 13

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, $N/p = 5$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .01$

Johansen's test		James' test
-	≥.0300	-
	≥.0275	
-	≥.0250	
	≥.0225	
-	≥.0200	---
--	≥.0175	--
--	≥.0150	--+
+++----===	≥.0125	=====----++++
++++-==	≥.0100	=++++
+++++=	≥.0075	+++
	≥.0050	

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 14

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, $N/p = 7$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .01$.

Johansen's test		James' test
	≥.0300	
	≥.0275	
	≥.0250	
	≥.0225	
	≥.0200	
	≥.0175	-
--	≥.0150	-
-----=	≥.0125	==------++
++++++-----==	≥.0100	==---+++++
+++++==	≥.0075	=-++++
+=	≥.0050	=+

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 15

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, $N/p = 9$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .01$

Johansen's test		James' test
	≥.0300	
	≥.0275	
	≥.0250	
	≥.0225	
	≥.0200	
	≥.0175	
--	≥.0150	-
+++-	≥.0125	---+++++
++++------=	≥.0100	=------+++
++++++--=	≥.0075	==--++++
-====	≥.0050	===-

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 16

Frequency Distributions of Estimated Type I Error Rates for the Beta Distribution, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .01$

Johansen's test		James' test
7555	≥.0300	555557
55	≥.0275	59
99975	≥.0250	59
55	≥.0225	5779
9777555	≥.0200	5555777
977755555	≥.0175	55577799
9999777775555	≥.0150	555557777999999
999999777755	≥.0125	555557777799999
9999999977777555555	≥.0100	55555777777999999999
999999777777755	≥.0075	7777779999
97	≥.0050	9

Note. The numbers 5, 7, and 9 indicate the three levels of N/p.

Figure 17

Frequency Distributions of Estimated Type I Error Rates for the Beta Distribution, $N/p = 5$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .01$

Johansen's test		James' test
---	≥.0300	=-----
--	≥.0275	-
-	≥.0250	-
-=	≥.0225	-
+-=	≥.0200	=-++
+---=	≥.0175	---
+++-	≥.0150	=++++
+=	≥.0125	=-+++
+++++=	≥.0100	==+++
+=	≥.0075	
	≥.0050	

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 18

Frequency Distributions of Estimated Type I Error Rates for the Beta Distribution, N/p = 7, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .01$

Johansen's test		James' test
-	≥.0300	-
	≥.0275	
-	≥.0250	
	≥.0225	--
---	≥.0200	=--
+--	≥.0175	--+
++--=	≥.0150	-+++
++-=	≥.0125	=---+
+++-=	≥.0100	===+++
++++===	≥.0075	=-++++
-	≥.0050	

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 19

Frequency Distributions of Estimated Type I Error Rates for the Beta Distribution, $N/p = 9$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .01$

Johansen's test		James' test
	≥.0300	
	≥.0275	-
---	≥.0250	-
	≥.0225	-
-	≥.0200	
-	≥.0175	--
+--=	≥.0150	==--++
+++--=	≥.0125	--+++
++++---=	≥.0100	=---+++++
+++===	≥.0075	===+
+	≥.0050	+

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 20

Frequency Distributions of Estimated Type I Error Rates for the Exponential Distribution, $N/p = 9$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .01$

Johansen's test		James' test
	≥.0400	-
--	≥.0375	-
-	≥.0350	-
-	≥.0325	-
	≥.0300	
-	≥.0275	-
-	≥.0250	-
--	≥.0225	--
	≥.0200	
	≥.0175	
--	≥.0150	--
-	≥.0125	-
-	≥.0100	-

Note. All results are for the negative relationship between sample sizes and v-c matrices.

Figure 21

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, Negative Condition, $\Sigma_2 = d^2\Sigma_1$, and $\alpha = .01$

Johansen's test		James' test
555	≥.0400	55
	≥.0375	5
	≥.0350	
	≥.0325	
	≥.0300	5
55	≥.0275	5
755	≥.0250	57
5	≥.0225	57
775	≥.0200	5
975	≥.0175	5777
97	≥.0150	559999
99999977775	≥.0125	5777999
999775	≥.0100	77779999
97	≥.0075	9

Note. The numbers 5, 7, and 9 indicate the three levels of N/p.

Figure 22

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, Negative Condition, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .01$

Johansen's test		James' test
	≥.0400	
	≥.0375	5
	≥.0350	
	≥.0325	
5	≥.0300	
	≥.0275	
5	≥.0250	
	≥.0225	
5	≥.0200	555
55	≥.0175	557
997755	≥.0150	5579
9777775555	≥.0125	5555777777999
999999777775	≥.0100	77799999
99	≥.0075	799
9	≥.0050	9

Note. The numbers 5, 7, and 9 indicate the three levels of N/p.

APPENDIX B

FREQUENCY DISTRIBUTIONS OF t WHEN $\alpha = .10$

Figure 23

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .10$

Johansen's test		James' test
5	≥.160	
5	≥.155	5
	≥.150	
5	≥.145	
55	≥.140	5
555	≥.135	5
9	≥.130	5
7755	≥.125	
99977555555	≥.120	559
999777775	≥.115	555577999
999977775555555	≥.110	5555555555777777999
999777775	≥.105	55777777999999
9999997777777555	≥.100	5555777777779999999
999999775	≥.095	555577777999999
997775	≥.090	7
99	≥.085	779999
	≥.080	

Note. The numbers 5, 7, and 9 indicate the three levels of N/p.

Figure 24

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, $N/p = 5$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .10$

Johansen's test		James' test
-	≥.160	
-	≥.155	-
	≥.150	
-	≥.145	
--	≥.140	-
---	≥.135	-
	≥.130	-
-=	≥.125	
+++--=	≥.120	--
=	≥.115	=--+
++++-==	≥.110	==-----++++
+	≥.105	=+
+++	≥.100	=+++
+	≥.095	=+++
=	≥.090	
	≥.085	
	≥.080	

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 25

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, $N/p = 7$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .10$

Johansen's test		James' test
	≥.160	
	≥.155	
	≥.150	
	≥.145	
	≥.140	
	≥.135	
	≥.130	
--	≥.125	
--	≥.120	
+---=	≥.115	--
++--	≥.110	=---++
+--==	≥.105	==---+
+++++==	≥.100	=---++++
+-	≥.095	=-+++
++=	≥.090	+
	≥.085	=+
	≥.080	

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 26

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, $N/p = 9$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .10$

Johansen's test		James' test
	≥.160	
	≥.155	
	≥.150	
	≥.145	
	≥.140	
	≥.135	
-	≥.130	
	≥.125	
---	≥.120	-
---	≥.115	---
-===	≥.110	=--
++-	≥.105	==--++
++++-=	≥.100	=--++++
++++-=	≥.095	=-++++
+-	≥.090	
+=	≥.085	=-++
	≥.080	

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 27

Frequency Distributions of Estimated Type I Error Rates for the Beta Distribution, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .10$

Johansen's test		James' test
755555	≥.160	555
75	≥.155	
75	≥.150	57
	≥.145	7
9755	≥.140	5
5555	≥.135	5579
9977555	≥.130	555
9777755	≥.125	555779
9999975555	≥.120	555577779999
9999777777555	≥.115	555555777777999
999977755	≥.110	5577999999
999977	≥.105	557777799999
9997777755	≥.100	577999
9999	≥.095	5777799999
9977	≥.090	5799
75	≥.085	
	≥.080	7

Note. The numbers 5, 7, and 9 indicate the three levels of N/p.

Figure 28

Frequency Distributions of Estimated Type I Error Rates for the Beta Distribution, $N/p = 5$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .10$

Johansen's test		James' test
-----	≥.160	---
-	≥.155	
-	≥.150	-
	≥.145	
+-	≥.140	-
+-==	≥.135	--
--=	≥.130	=-+
+-	≥.125	=-+
++++	≥.120	=-++
+==	≥.115	--++++
++	≥.110	=+
	≥.105	=+
++	≥.100	+
	≥.095	+
	≥.090	=
=	≥.085	
	≥.080	

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 29

Frequency Distributions of Estimated Type I Error Rates for the Beta Distribution, $N/p = 7$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .10$

Johansen's test		James' test
-	≥.160	
-	≥.155	
-	≥.150	-
	≥.145	-
-	≥.140	
	≥.135	-
--	≥.130	
+---	≥.125	-+
=	≥.120	----
+++-==	≥.115	====--+
++=	≥.110	++
+-	≥.105	=-+++
+++-=	≥.100	=+
	≥.095	=-++
+=	≥.090	+
+	≥.085	
	≥.080	+

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 30

Frequency Distributions of Estimated Type I Error Rates for the Beta Distribution, N/p = 9, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .10$

Johansen's test		James' test
	≥.160	
	≥.155	
	≥.150	
	≥.145	
-	≥.140	
	≥.135	-
--	≥.130	
-	≥.125	-
+----	≥.120	--++
++--	≥.115	--+
+-==	≥.110	=-----
+++=	≥.105	=++++
++-	≥.100	=++
+===	≥.095	===-+
++	≥.090	++
	≥.085	
	≥.080	

Note. A minus sign indicates a ↑ for a negative condition, a positive sign a ↑ for a positive condition, and an equal sign a ↑ for an equal-sample-size condition.

Figure 31

Frequency Distributions of Estimated Type I Error Rates for the Exponential Distribution, $N/p = 9$, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .10$

Johansen's test		James' test
--	≥.170	
-	≥.165	
--	≥.160	--
	≥.155	---
-	≥.150	
-	≥.145	-
-	≥.140	
	≥.135	--
-	≥.130	
	≥.125	-
-	≥.120	
	≥.115	-
-	≥.110	-
-	≥.105	-

Note. All results are for the negative relationship between sample sizes and v-c matrices.

Figure 32

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, Negative Condition, $\Sigma_2 = d^2\Sigma_1$, and $\alpha = .10$

Johansen's test		James' test
775555555	≥.160	5555
	≥.155	
7	≥.150	5
955	≥.145	5
97	≥.140	
775	≥.135	55777
95	≥.130	59
	≥.125	59
997	≥.120	77
9977	≥.115	9
995	≥.110	55779999
97	≥.105	779
997	≥.100	77999
7	≥.095	79

Note. The numbers 5, 7, and 9 indicate the three levels of N/p.

Figure 33

Frequency Distributions of Estimated Type I Error Rates for the Normal Distribution, Negative Condition, $\Sigma_2 \neq d^2\Sigma_1$, and $\alpha = .10$

Johansen's test		James' test
5	≥.160	
5	≥.155	5
	≥.150	
5	≥.145	
55	≥.140	5
555	≥.135	5
9	≥.130	5
775	≥.125	
9997755	≥.120	559
999777	≥.115	5577999
9775	≥.110	555577799
977	≥.105	77799
9	≥.100	77799
97	≥.095	79
9	≥.090	
	≥.085	9

Note. The numbers 5, 7, and 9 indicate the three levels of N/p.

APPENDIX C

SUMMARY ANOVA TABLES FOR $\alpha = .01$ AND $\alpha = .10$

Table 28

Summary ANOVA Table: Analysis 1 and $\alpha = .01$

Source of Variation	Degrees of Freedom	F Value	PR > F
Test (T)	1	4.96	.0267*
Distribution (D)	1	132.82	.0001*
Relationship (R)	1	268.64	.0001*
p	2	12.52	.0001*
N/p	2	86.72	.0001*
$n_1:n_2$	1	62.71	.0001*
d	1	91.36	.0001*
T x D	1	0.07	.7860
T x R	1	0.69	.4079
T x p	2	5.85	.0032*
T x N/p	2	0.31	.7353
T x $n_1:n_2$	1	0.15	.6992
T x d	1	0.01	.9148
D x R	1	53.45	.0001*
D x p	2	6.57	.0016*
D x N/p	2	7.06	.0010*
D x $n_1:n_2$	1	11.36	.0008*
D x d	1	98.16	.0001*
R x p	2	12.68	.0001*
R x N/p	2	39.05	.0001*
R x $n_1:n_2$	1	20.36	.0001*
R x d	1	68.07	.0001*
p x N/p	4	0.76	.5552
p x $n_1:n_2$	2	2.95	.0538
p x d	2	11.97	.0001*
N/p x $n_1:n_2$	2	17.28	.0001*
N/p x d	2	13.35	.0001*
$n_1:n_2$ x d	1	4.56	.0334*
Error	308		

Note. When testing the two factors, R and $n_1:n_2$, or any interaction involving either of these two factors, the data under equal-sample-size condition are not taken into consideration by GLM in SAS. Thus, the degrees of freedom have to be reduced by one or adjusted accordingly.

* $p < .05$.

Table 29

Summary ANOVA Table: Analysis 2 and $\alpha = .01$

Source of Variation	Degrees of Freedom	F Value	PR > F
Test (T)	1	0.27	.6043
Distribution (D)	2	243.79	.0001*
p	2	29.92	.0001*
$n_1:n_2$	1	31.19	.0001*
d	1	318.60	.0001*
T x D	2	0.05	.9516
T x p	2	0.61	.5502
T x $n_1:n_2$	1	0.17	.6783
T x d	1	0.02	.8763
D x p	4	2.46	.0587
D x $n_1:n_2$	2	12.57	.0001*
D x d	2	59.73	.0001*
p x $n_1:n_2$	2	0.86	.4288
p x d	2	13.74	.0001*
$n_1:n_2$ x d	1	2.45	.1244
Error	45		

* $p < .05$.

Table 30

Summary ANOVA Table: Analysis 3 and $\alpha = .01$

Source of Variation	Degrees of Freedom	F Value	PR > F
Test (T)	1	0.40	.5291
p	2	7.99	.0006*
N/p	2	126.19	.0001*
$n_1:n_2$	1	59.47	.0001*
d	1	78.53	.0001*
Type of heterogeneity of v-c matrices (TH)	1	66.99	.0001*
T x p	2	4.23	.0170*
T x N/p	2	0.07	.9350
T x $n_1:n_2$	1	0.06	.8046
T x d	1	0.01	.9283
T x TH	1	0.00	1.0000
p x N/p	4	2.93	.0242*
p x $n_1:n_2$	2	3.08	.0498*
p x d	2	2.09	.1287
p x TH	2	0.44	.6483
N/p x $n_1:n_2$	2	34.25	.0001*
N/p x d	2	19.98	.0001*
N/p x TH	2	17.49	.0001*
$n_1:n_2$ x d	1	14.35	.0002*
$n_1:n_2$ x TH	1	13.34	.0004*
d x TH	1	36.24	.0001*
Error	109		

* $p < .05$.

Table 31

Summary ANOVA Table: Analysis 1 and $\alpha = .10$

Source of Variation	Degrees of Freedom	F Value	PR > F
Test (T)	1	35.48	.0001*
Distribution (D)	1	96.72	.0001*
Relationship (R)	1	393.51	.0001*
p	2	13.41	.0001*
N/p	2	121.05	.0001*
$n_1:n_2$	1	124.51	.0001*
d	1	90.45	.0001*
T x D	1	0.02	.8766
T x R	1	11.44	.0008*
T x p	2	11.91	.0001*
T x N/p	2	4.71	.0096*
T x $n_1:n_2$	1	2.26	.1335
T x d	1	1.54	.2160
D x R	1	14.30	.0002*
D x p	2	2.47	.0865
D x N/p	2	3.86	.0220*
D x $n_1:n_2$	1	11.24	.0009*
D x d	1	30.15	.0001*
R x p	2	0.33	.7207
R x N/p	2	22.21	.0001*
R x $n_1:n_2$	1	31.52	.0001*
R x d	1	16.86	.0001*
p x N/p	4	1.07	.3737
p x $n_1:n_2$	2	1.72	.1812
p x d	2	3.30	.0381*
N/p x $n_1:n_2$	2	22.19	.0001*
N/p x d	2	2.93	.0547
$n_1:n_2$ x d	1	1.00	.3173
Error	308		

Note. When testing the two factors, R and $n_1:n_2$, or any interaction involving either of these two factors, the data under equal-sample-size condition are not taken into consideration by GLM in SAS. Thus, the degrees of freedom have to be reduced by one or adjusted accordingly.

* $p < .05$.

Table 32

Summary ANOVA Table: Analysis 2 and $\alpha = .10$

Source of Variation	Degrees of Freedom	F Value	PR > F
Test (T)	1	9.67	.0032*
Distribution (D)	2	113.53	.0001*
p	2	5.12	.0099*
$n_1:n_2$	1	41.32	.0001*
d	1	81.98	.0001*
T x D	2	0.03	.9716
T x p	2	1.23	.3029
T x $n_1:n_2$	1	0.53	.4694
T x d	1	0.47	.4978
D x p	4	3.90	.0084*
D x $n_1:n_2$	2	5.07	.0103*
D x d	2	15.76	.0001*
p x $n_1:n_2$	2	0.99	.3785
p x d	2	7.62	.0014*
$n_1:n_2$ x d	1	0.41	.5271
Error	45		

* $p < .05$.

Table 33

Summary ANOVA Table: Analysis 3 and $\alpha = .10$

Source of Variation	Degrees of Freedom	F Value	PR > F
Test (T)	1	47.62	.0001*
p	2	9.99	.0001*
N/p	2	155.27	.0001*
$n_1:n_2$	1	105.89	.0001*
d	1	118.12	.0001*
Type of heterogeneity of v-c matrices (TH)	1	84.30	.0001*
T x p	2	10.52	.0001*
T x N/p	2	2.79	.0656
T x $n_1:n_2$	1	2.02	.1581
T x d	1	3.05	.0836
T x TH	1	2.91	.0908
p x N/p	4	1.29	.2778
p x $n_1:n_2$	2	1.53	.2204
p x d	2	1.46	.2367
p x TH	2	2.84	.0626
N/p x $n_1:n_2$	2	23.53	.0001*
N/p x d	2	10.68	.0001*
N/p x TH	2	23.52	.0001*
$n_1:n_2$ x d	1	8.70	.0039*
$n_1:n_2$ x TH	1	15.02	.0002*
d x TH	1	31.65	.0001*
Error	109		

* $p < .05$.

REFERENCES

Algina, J., & Oshima, T. C. (1990a). Robustness of the independent sample Hotelling's T^2 to variance-covariance heteroscedasticity when sample sizes are unequal or in small ratios. Psychological Bulletin, 108, 308-313.

Algina, J., & Oshima, T. C. (1990b). Robustness of Yao's, James', and Johansen's tests under variance-covariance heteroscedasticity and non-normality. Manuscript submitted for publication.

Algina, J., & Tang, K. L. (1988). Type I error rates for Yao's and James' tests of equality of mean vectors under variance-covariance heteroscedasticity. Journal of Educational Statistics, 13, 281-290.

Bradley, J. V. (1978). Robustness? British Journal of Mathematical and Statistical Psychology, 31, 144-152.

Clinch, J. J., & Keselman, H. J. (1982). Parametric alternatives to the analysis of variance. Journal of Educational Statistics, 7, 207-214.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. The American Statistician, 35, 124-129.

Everitt, B. S. (1979). A Monte Carlo investigation of the robustness of Hotelling's one- and two-sample T^2 tests. Journal of the American Statistical Association, 74, 48-51.

Fisher, R. A. (1935). The fiducial argument in statistical inference. Annals of Eugenics, 6, 391-398.

Fisher, R. A. (1938). Statistical tables for biological, agricultural, and medical research (1st ed.). Edinburgh, Scotland: Oliver and Boyd.

Fisher, R. A. (1939). The comparison of samples with possible unequal variances. Annals of Eugenics, 9, 174-180.

Fligner, M. A., & Policello, G. E. (1981). Robust rank procedures for the Behrens-Fisher problem. Journal of the American Statistical Association, 76, 162-168

Hakstian, A. R., Roed, J. C., & Lind, J. C. (1979). Two-sample T^2 procedure and the assumption of homogeneous covariance matrices. Psychological Bulletin, 86, 1255-1263.

Holloway, L. N., & Dunn, O. J. (1967). The robustness of Hotelling's T^2. Journal of the American Statistical Association, 62, 124-136.

Hopkins, J. W., & Clay, P. P. F. (1963). Some empirical distributions of bivariate T^2 and homoscedasticity criterion M under unequal variance and leptokurtosis. Journal of the American Statistical Association, 58, 1048-1053.

Hotelling, H. (1931). The generalization of Student's ratio. Annals of Mathematical Statistics, 2, 360-378.

Ito, K. (1969). On the effect of heteroscedasticity and nonnormality upon some multivariate test procedures. In P. R. Krishnaiah (Ed.), Multivariate Analysis-II (pp.87-120). New York: Academic Press.

Ito, K., & Schull, W. J. (1964). On the robustness of the T^2 test in multivariate analysis of variance when variance-covariance matrices are not equal. Biometrika, 38, 324-329.

James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. Biometrika, 38, 324-329.

James, G. S. (1954). Tests of linear hypothesis in univariate and multivariate analysis when the ratios of the population variances are unknown. Biometrika, 41, 19-43.

Johansen, S. (1980). The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. Biometrika, 67, 85-92.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics, 18, 50-60.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin, 105, 156-166.

Nath, R., & Duran, B. S. (1983). A robust test in the multivariate two-sample location problem. American Journal of Mathematical and Management Sciences, 3, 225-249.

Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. Journal of the American Statistical Association, 69, 894-908.

Pavur, R., & Nath, R. (1989). Power and Type I error rates for rank-score MANOVA techniques. Multivariate Behavioral Research, 24, 477-501.

Pearson, E. S., & Please, N. W. (1975). Relation between the shape of population distribution and the robustness of four simple test statistics. Biometrika, 62, 223-241.

Puri, M. L., & Sen, P. K. (1971). Nonparametric methods in multivariate analysis. New York: John Wiley.

SAS Institute Inc. (1979). SAS user's guide, 1979 edition. Raleigh, NC: Author.

Tang, K. L. (1989). Robustness of four multivariate tests under variance-covariance heteroscedasticity. Unpublished doctoral dissertation, University of Florida. (University Microfilms No. 90-28,574)

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. Biometrika, 29, 350-362.

Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. Biometrika, 34, 28-35.

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. Biometrika, 38, 330-336.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics, 1, 80-83.

Yao, Y. (1965). An approximate degrees of freedom solution to the multivariate Behrens-Fisher problem. Biometrika, 52, 139-147.

Zwick, R. (1986). Rank and normal scores alternatives to Hotelling's T^2. Multivariate Behavioral Research, 21, 169-186.

BIOGRAPHICAL SKETCH

Wen-ying Lin was born in Taipei, Taiwan, on August 20, 1962. She received her B.Ed. from National Taiwan College of Education, Changhua, Taiwan, in June, 1984. She then attended Southern Illinois University at Carbondale, Illinois, where she received her M.A. in English as a Foreign Language in August, 1986.

In the fall of 1986, she began studying for the Ph.D. in educational testing, research, and evaluation at the University of Florida. She will graduate with the Ph.D. in May, 1991.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James J. Algina, Chair
Professor of Foundations
of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Linda M. Crocker
Professor of Foundations
of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

M. David Miller
Assistant Professor of
Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert R. Sherman
Professor of Foundations
of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Walter R. Cunningham
Professor of Psychology

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May, 1991

Chairman, Foundations of Education

Dean, College of Education

Dean, Graduate School